GAIN OF FUNCTION MUTATIONS IN ATP-DEPENDENT TRANSPOSITION PROTEINS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation of U.S. Application No. 09/027,169, filed February 20, 1998, which claims priority to provisional application no. 60/037,955 filed on February 20, 1997, the teachings of which are hereby incorporated herein in their entirety by reference.

BACKGROUND OF THE INVENTION

The invention is specifically directed to efficient, random, simple insertion of a transposon or derivative transposable element into DNA *in vivo* or *in vitro*. The invention is particularly directed to mutations in ATP-utilizing regulatory transposition proteins that permit insertion with less target-site specificity than wild-type. The invention encompasses gain-of-function mutations in TnsC, an ATP-utilizing regulatory transposition protein that activates the bacterial transposon Tn7. Such mutations enable the insertion of a Tn7 transposon or derivative transposable element in a non-specific manner into a given DNA segment. Insertion can be effected in plasmid and cosmid libraries, cDNA libraries, PCR products, bacterial artificial chromosomes, yeast artificial chromosomes, mammalian artificial chromosomes, genomic DNAs, and the like. Such insertion is useful in DNA sequencing methods, for genetic analysis by insertional mutagenesis, and alteration of gene expression by insertion of a given genetic sequence.

Description of the Background Art

Transposable elements are discrete segments of DNA capable of mobilizing nonhomologously from one genetic location to another, that typically carry sequence information important for two main functions that confer the ability to mobilize. They encode the proteins necessary to carry out the catalytic activity associated with transposition, and contain the cis-acting sequences, located at the transposon termini, that act as substrates for these proteins. The same proteins can participate in the selection of the target site for insertion.

The selection of a new insertion site is usually not a random process; instead, many transposons show characteristic preferences for certain types of target sites. One broad characteristic that differentiates the wide variety of transposable elements known is the nature of the target site selectivity (1). A component of this selectivity can be the target sequence itself. The bacterial transposon Tn10 preferentially selects a relatively highly conserved 9 bp motif as the predominant site for transposon insertion and less often selects other more distantly related sites *in vivo* (2). The Tc1 and Tc3 mariner elements of *C. elegans* insert preferentially at a TA dinucleotide such that each end of the element is flanked by a TA duplication (3) (4) (5). A lower specificity consensus sequence, N-Y-G/C-R-N has been determined from populations of both *in vivo* and *in vitro* insertions for the bacteriophage Mu (7). In contrast to these elements, the bacterial transposon Tn5 exhibits markedly lower insertion site specificity, although some isolated "hotspots" have been detected (8).

Another selection mechanism relies on structural features or presence of cellular protein complexes at the target sites. The yeast transposon Ty3 preferentially inserts into the promoters of genes transcribed by RNA polymerase III, responding to signals from cellular proteins TFIIIB and TFIIIC (9).

Understanding how these factors modulate transposase activity to impose target site preferences will lend insight into the spread of transposons and viruses, and may suggest ways to manipulate those target preferences. The bacterial transposon Tn7 is distinctive in

that it uses several element-encoded accessory proteins to evaluate potential target DNAs for positive and negative features, and to select a target site (1). Tn7 encodes five genes whose protein products mediate its transposition (10) (11).

Two of the proteins, TnsA and TnsB, constitute the transposase activity, collaborating to execute the catalytic steps of strand breakage and joining (12). The activity of this transposase is modulated by the remaining proteins, TnsC, TnsD, and TnsE, and also by the nature of the target DNA.

TnsC, TnsD, and TnsE interact with the target DNA to modulate the activity of the transposase via two distinct pathways. TnsABC + TnsD directs transposition to *attTn7,* a discrete site on the *E. coli* chromosome, at a high frequency, and to other loosely related "pseudo att" sites at low frequency (13). The alternative combination TnsABC + E directs transposition to many unrelated non-*attTn7* sites in the chromosome at low frequency (13) (10) (11) and preferentially to conjugating plasmids (14). Thus, *attTn7* and conjugable plasmids contain positive signals that recruit the transposon to these target DNAs. The alternative target site selection mechanisms enable Tn7 to inspect a variety of potential target sites in the cell and select those most likely to ensure its survival.

The Tn7 transposition machinery can also recognize and avoid targets that are unfavorable for insertion. Tn7 transposition occurs only once into a given target molecule; repeated transposition events into the same target are specifically inhibited (15) (16). Therefore, a pre-existing copy of Tn7 in a potential target DNA generates a negative signal which renders that target "immune" to further insertion. The negative target signal affects both TnsD- and TnsE-activated transposition reactions and is dominant to any positive signals present on a potential target molecule (16). Several other transposons, such as Mu and members of the Tn3 family, also display this form of negative target regulation (17) (18) (19) (7).

Target selection could be an early or late event in the course of a transposition reaction. For example, a transposon could constitutively excise from its donor position, and the excised transposon could then be captured at different frequencies by different types of target molecules. Tn10 appears to follow this course of events *in vitro*, excising

from its donor position before any interactions with target DNA occur (20) (21). Alternatively, the process of transposon excision could itself be dependent on the identification of a favorable target site. Tn7 transposition shows an early dependence on target DNA signals *in vitro*: neither transposition intermediates nor insertion products are seen in the absence of an *attTn7* target (22). Thus, the nature of the target DNA appears to regulate the initiation of Tn7 transposition *in vitro*.

An important question is how positive and negative target signals are communicated to the Tn7 transposase. Reconstitution of the TnsABC + TnsD reaction *in vitro* has provided a useful tool for detailed dissection of Tn7 transposition (22) (23). This reaction has been instrumental in delineating the role of each of the individual proteins play in target site selection. Dissection of the TnsABC+D reaction *in vitro* has implicated TnsC as a pivotal connector between the TnsAB transposase and the target DNA. TnsC is an ATP-dependent DNA-binding protein with no known sequence specificity (24). However, TnsC can respond to signals from *attTn7* via an interaction with the site-specific DNA-binding protein TnsD. In a standard *in vitro* transposition reaction TnsD is required for transposition to the *attTn7* site on a target DNA molecule. This site-specific insertion process is tightly regulated by TnsC, but does not occur in the absence of TnsD. Additional evidence for a TnsC-TnsD interaction comes from DNA protection and band shift analysis with *attTn7* DNA (23). Direct interaction between TnsC and the TnsAB transposase has also recently been observed (25) (26).

Therefore, TnsC may serve as a "connector" or "matchmaker" between the transposase and the TnsD+*attTn7* target complex (23) (27). This connection is not constitutive, but instead appears to be regulated by the ATP state of TnsC. Only the ATP-bound form of TnsC is competent to interact with target DNAs and activate the TnsA+B transposase; the ADP-bound form of TnsC has neither of these activities and cannot participate in Tn7 transposition (24) (23). TnsC hydrolyzes ATP at a modest rate (25), and therefore can switch from an active to an inactive state. The modulation of the ATP state of TnsC may be a central mechanism for regulating Tn7 transposition.

The possibility that TnsC regulates the connection between the TnsA+B transposase and the target site prompted the inventor to predict that TnsC mutants can be isolated that would constitutively activate Tn7 transposition.

TnsC therefore became an excellent candidate for mutagenesis, to search for a gain of function protein capable of circumventing the requirement for targeting proteins. The inventor therefore identified gain-of-function TnsC mutants which can activate the TnsA+B transposase in the absence of TnsD or TnsE. They have characterized the ability of these mutants to promote insertions into various targets, and to respond to regulatory signals on those targets.

One class of TnsC mutants activates transposition in a way that is still sensitive to target signals, whereas a second class of TnsC mutants activates transposition in a way that appears to bypass target signals. As had been observed *in vitro*, the critical communication between the transposon and the target DNA appears to be an early event in the Tn7 reaction pathway *in vivo*, preceding the double-strand breaks at the transposon ends that initiate transposition.

A particular mutant isolated from the random mutagenesis is $TnsC^{A225V}$, a mutant capable of an impressive activation of Tn7 transposition in the absence of TnsD (25). The single amino acid substitution made to generate $TnsC^{A225V}$ has altered the protein such that it no longer requires an interaction with the target-associated TnsD, enabling it to activate transposition to a variety of target molecules very efficiently (25) (26). The inventor concluded that $TnsC^{A225V}$ could promote transposition to target DNAs with low specificity based on results where transposition driven by the $TnsABC^{A225V}$ machinery was directed to either F plasmids containing an *attTn7* site, F plasmids lacking an *attTn7* site, or the *E. coli* chromosome with no apparent preference.

DNA Sequencing

Sequencing DNA fragments cloned into vectors requires provision of priming sites at distributed locations within the fragment of interest, if the fragment is larger than the sequence run length (amount of sequence that can be determined from a single sequencing

reaction). At present there are three commonly used methods of providing these priming sites:

A) Design of a new primer from sequence determined in a previous run from vector-encoded primer or other previously determined primer (prime and run, primer walking)

B) Random fragmentation and recloning of smaller pieces, followed by determination of the sequence of the smaller pieces from vector-encoded (universal) priming sites, followed by sequence assembly by overlap of sequence (random shotgun sequencing).

C) Deletion of variable amounts of the fragment of interest from an end adjacent to the vector, to bring undetermined fragment sequence close enough to the vector-encoded (universal) primer to allow sequence determination.

All of these methods have disadvantages.

Method A is time-consuming and expensive because of the delay involved in design of new primers and their cost. Moreover, if the fragment contains DNA repeats longer than the sequence run, it may be impossible to design a unique new primer; sequence runs made with primers within the repeat sequence will display two or more sequences that cannot be disentangled.

Method B requires recloning; random fragmentation is difficult to achieve because fragments that are efficiently clonable (restriction enzyme digestion) do not have ends randomly distributed (Adams, M.D., Fields, C. and Ventor, J. C. editors Automated DNA Sequencing and Analysis *Academic Press* 1994; Chapter 6, Bodenteich, K. et al.), and fragmentation methods that provide randomly distributed ends (shearing, sonication) do not provide DNA ends that are efficiently clonable (with 5' phosphate and 3' OH moieties). Sequence assembly of is also difficult or impossible when two or more repetitive sequences longer than the sequence run are present in the starting fragment.

Method C depends on providing randomly distributed end points for enzymatically - determined deletions. There are many methods for making such deletions (especially those involving exonuclease digestions, typically Exonuclease III), none of which provide

entirely random endpoints and which depend on the presence of unique suitable restriction enzyme sites at one or both ends of the cloned fragment. However, because the deletion series in principle allows construction of a map (of nested remaining fragment lengths in deletion derivatives) that is independent of the sequence itself, this method can allow repetitive sequence longer than the sequence run to be located within the fragment at appropriate locations.

A method for introduction of universal priming sites at randomly distributed locations within a fragment of interest is therefore a useful advance in sequencing technology.

Transposition and the sequencing problem.

Previous efforts have been made to provide distributed priming sites by means of transposable elements. These methods have fallen short of this goal in three ways: first, the transposable elements have not provided a sufficiently random distribution of priming sites; second, the transposition method (carrying out transposition *in vivo*, followed by recovery of the targeted DNA and repurification) has been time-consuming and laborious; third, the Systems have been prone to produce undesired products. These undesired products include but are not limited to: a) cointegrates (replicon fusions) between the donor of the transposon and the target plasmid; b) insertions in which the two ends of the transposon act at different positions (leading to deletion of the intervening target); c) insertions of multiple copies of the transposon into the target, so that priming from one end of the transposon yields two superimposed sequences. The method has been laborious in two ways: the majority of insertions have been into chromosomal DNA of the host, and even for those insertions into the plasmid the recovery method has entailed loss of independence of insertions. *in vitro* methods of insertion have suffered from both the non-random location of insertion sites and the undesired products, and also from poor efficiency, so that it has been impractical to obtain large numbers of insertions into the target of interest without excessive labor.

Increasing interest in large scale sequencing projects and a concomitant search for highly efficient *in vitro* mutagenesis methods has promoted the adaptation of several *in*

vitro transposon systems as tools to study genomes. An *in vivo* reaction for the bacterial transposon Tn3 has been used to efficiently sequence plasmid inserts of variable lengths; however, only approximately 37% of the nucleotides were found to be capable of serving as sites for insertion (Davies, 1995 #419). A similar, more random system has been developed for yeast retrotransposon Ty1, employing synthetic transposons with U3 ends as substrates and Ty1 virus-like particles supplying transposition functions (28) to sequence plasmids with yeast and human DNA inserts. A disadvantage to this method is the requirement for the cumbersome preparation of VLPs. *In vitro* transposition with an MLV integrase system has been utilized as a tool to dissect some of the mysteries of chromatin packaging (29) (30) (31) and as a tool for functional genetic footprinting (32). However, the MLV insertions do not appear to be completely random. An object of the invention therefore is to provide a transposon and transposition reaction with more random target site specificity. Therefore, the inventor examined the target site selectivity of the $TnsC^{A225V}$ machinery *in vitro* and explored the viability of this reaction as an effective tool for random insertional mutagenesis.

BRIEF SUMMARY OF THE INVENTION

Accordingly, a general object of the invention is to provide a transposable system that achieves efficient, simple, non-specific or random insertion into any given DNA segment.

A further object of the invention is to provide a transposable system that achieves efficient random insertional mutagenesis *via* simple insertion.

Therefore, a specific object of the invention is to provide a transposable system that achieves efficient target site specificity that is reduced from wild-type and preferably random, *via* simple insertion.

A more particular object of the invention is to provide a transposon containing a mutation in a transposon-derived protein that allows efficient, simple insertion and target site selectivity that is reduced from the wild-type, and preferably random.

A more particular object of the invention is to provide a transposable system with a mutation in a transposon-derived ATP-utilizing regulatory protein. The mutation allows the efficient, simple, non-specific or random insertion of the transposable element into a DNA segment or at least provides reduced target site specificity from the wild-type.

A preferred object of the invention is to provide a Tn7 transposable system that achieves simple, efficient, non-specific or random insertion into a given DNA segment, or at least reduced target site specificity compared to the wild-type Tn7.

A preferred object of the invention is to provide a mutation in the Tn7 transposon that confers efficient, simple, non-specific insertion into a given DNA segment, or at least reduced target site specificity compared to the wild-type Tn7.

A preferred object of the invention is to provide a Tn7 transposable system with a mutation in the TnsC protein encoded in the Tn7 transposon, which mutation allows efficient, simple insertion with reduced target site specificity compared to the wild-type, and preferably allows non-specific insertion into a DNA segment.

Objects of the invention include methods for using the above compositions.

Accordingly, a general object of the invention is to provide a method for efficient, simple, random insertion of a transposable element into a given DNA segment.

A further object of the invention is to provide a method for efficient, simple, random insertional mutagenesis by a transposable element.

A specific object of the invention is to provide a method for efficient, simple, random transposition of a transposable element into a DNA segment, or in which the specificity of transposition is reduced compared to wild-type.

A more particular object of the invention is to provide a method for efficient, simple, random transposition of a transposable element into a DNA segment in which the specificity of transposition is reduced compared to the wild-type by using a transposable system containing a mutation that confers efficient, simple insertion with reduced target site specificity compared to the wild-type, and preferably random insertion.

A more particular object of the invention is to provide a method for efficient, simple, random transposition of a transposable element into a DNA segment or in which

the specificity of transposition is reduced compared to wild-type, by using a transposable system with a mutation in an ATP-utilizing regulatory protein, the mutation allowing the efficient, simple, non-specific insertion of the transposable element into a DNA segment or at least providing for reduced target site specificity compared to the wild-type.

A preferred object of the invention is to provide a method for efficient, simple transposition of a transposable element into a DNA segment in which the specificity of transposition is reduced compared to wild-type, or is preferably random, by providing a Tn7 transposable system that is capable of non-specific insertion into a DNA segment, or at least reduced target site specificity compared to the wild-type Tn7.

A further object of the invention is to provide a method for efficient, simple transposition of a transposable element transposon into a DNA segment in which specificity of transposition is reduced compared to wild-type or is preferably random by providing a Tn7 mutation that confers the efficiency, ability to make a simple insertion, and the randomness or reduced specificity.

A further object of the invention is to provide a method for efficient, simple, random transposition of a transposable element into a DNA segment, or in which the specificity of transposition is reduced compared to the wild-type, by providing a mutation in the TnsC protein encoded in the Tn7 transposon, the mutation allowing a reduction in target site specificity compared to the wild-type and preferably allowing non-specific or random insertion of the Tn7 transposable element into a DNA segment.

A further object of the invention is to provide a method for DNA sequencing using a transposable system to introduce priming sites at randomly-distributed locations within a fragment of interest where the fragment is larger than the sequence run length.

A preferred object of the invention is to provide a method for DNA sequencing using a transposable system with a mutation that allows efficient and simple insertion and target site selectivity that is reduced from the wild-type and preferably random.

A preferred object of the invention is to provide a mutation in an ATP-utilizing regulatory protein. The mutation allows the efficient, simple, non-specific insertion of the

transposon into a DNA segment or at least provides reduced target site specificity over wild-type.

A highly preferred object of the invention is to provide a method for DNA sequencing using a Tn7 transposable system that allows efficient, simple, non-specific insertion into a DNA segment or at least reduced target site specificity compared to the wild-type Tn7.

A highly preferred object of the invention is to provide a method for DNA sequencing using a Tn7 transposable system with a mutation in the TnsC protein, the mutation allowing efficient, simple insertion and a reduction in target site specificity compared to the wild-type and preferably allowing non-specific or random insertion of the Tn7 transposable element into the DNA segment.

A further object of the invention is to provide methods as described above that can be applied to any given DNA segment. These include, but are not limited to, plasmids, cellular genomes, including prokaryotic and eukaryotic, bacterial artificial chromosomes, yeast artificial chromosomes, and mammalian artificial chromosomes, and subsegments of any of these.

An object of the invention is to provide these methods *in vitro* or *in vivo*.

A further object of the invention is to provide kits for carrying out the above-described methods using the above-described transposons or parts thereof.

The inventor has accordingly developed a transposable system and methods that improve on *in vitro* and *in vivo* transmission methods previously described in that the methods are efficient for transposition, provide relatively random insertion, and almost all products recovered are simple insertions at a single site which thus provide useful information.

In a general embodiment of the invention, the invention is directed to a transposable system that achieves simple, efficient, random insertion into a given DNA segment.

In a further embodiment of the invention, the invention is directed to a transposable system that is capable of efficient random insertional mutagenesis, preferably by means of a simple insertion.

In a specific embodiment of the invention, the invention is directed to a transposable system with target site specificity that is reduced from the wild-type and preferably random, which allows simple and efficient insertion.

In a further specific embodiment of the invention, the invention is directed to a transposable system containing a mutation that allows target site specificity that is reduced from the wild-type and is preferably random.

In a preferred embodiment of the invention, the invention is directed to a transposable system with a mutation in an ATP-utilizing regulatory protein, the mutation allowing the efficient, simple, non-specific insertion of the transposon into a DNA segment or at least providing reduced target site specificity from the wild-type.

In a highly preferred embodiment of the invention, the invention is directed to a Tn7 transposable system that achieves efficient, simple, non-specific insertion into a given DNA segment, or at least reduced target site specificity compared to the wild-type Tn7.

In a highly preferred embodiment of the invention, the invention is directed to a mutation in a Tn7 transposon that confers the capability of efficient, simple, non-specific insertion into a DNA segment, or at least reduced target site specificity compared to the wild-type Tn7.

In a highly preferred embodiment of the invention, the invention is directed to a mutation in the TnsC protein encoded in the Tn7 transposon, the mutation allowing simple, efficient insertion and a reduction in target site specificity compared to the wild-type and preferably allowing non-specific or random insertion of the Tn7 transposition into a DNA segment.

In a specific disclosed embodiment of the invention, the invention is directed to a Tn7 mutant designated $TnsC^{A225V}$, which is a mutant having an alanine to valine substitution at amino acid number 225 in the TnsC gene.

The invention also embodies methods for using all of the above compositions. Methods are directed to transposition or insertion of the transposable elements described above.

Accordingly, in one embodiment, the invention provides generally for efficient, simple, random insertion of a transposon into a given DNA segment, or at least insertion with reduced specificity compared to the wild-type.

In a further embodiment of the invention, the invention is directed to methods for insertional mutagenesis using a transposable system that is capable of efficient, simple, random insertion or at least insertion with reduced specificity compared to wild-type.

In a further embodiment of the invention, the invention is directed to methods for insertion of a transposable element into a DNA segment in which target site specificity is reduced from wild-type and is preferably random, where insertion is efficient and simple.

In a further embodiment of the invention, the invention is directed to methods for insertion of a transposable element into a DNA segment, by providing a transposable element containing a mutation that allows efficient and simple insertion and target site specificity that is reduced from the wild-type and is preferably random.

In a preferred embodiment of the invention, the invention is directed to methods for inserting a transposable element into a DNA segment by providing a transposable system with a mutation in an ATP-utilizing regulatory protein, the mutation allowing simple, efficient, and non-specific insertion of the transposon into a DNA segment, or at least providing reduced target site specificity from the wild-type.

In a highly preferred embodiment of the invention, the invention is directed to methods for inserting a transposable element into a DNA segment by providing a Tn7 transposable system allowing efficient, simple, non-specific insertion into a given DNA segment or at least reduced target site specificity compared to the wild-type Tn7.

In a highly preferred embodiment of the invention, the invention is directed to a Tn7 transposable system with a mutation that allows simple, efficient, and non-specific insertion of a transposable element into a DNA segment or at least provides reduced target site specificity from the wild-type Tn7.

In a highly preferred embodiment of the invention, the invention is directed to methods for inserting a transposable element into a DNA segment by providing a Tn7 transposable system with a mutation in the TnsC protein, the mutation allowing efficient

and simple insertion and a reduction in target site specificity compared to the wild-type and preferably allowing non-specific or random insertion of the Tn7 transposition into a DNA segment.

In a specific disclosed embodiment of the invention, the invention is directed to methods for inserting a transposable element into a DNA segment, by providing the Tn7 mutant $TnsC^{A225V}$.

The invention also provides kits for performing the above-described methods and the methods further described herein. In a preferred embodiment, a kit is supplied whose components comprise a mutant ATP-utilizing regulatory protein derived from a transposon, the mutation allowing efficient, simple, non-specific insertion of the transposon into a given DNA segment. The kit also provides a transposable element which can be found as part of a larger DNA segment; for example, a donor plasmid. The kit can further comprise a buffer compatible with insertion of the transposable element. The kit can further comprise a control target sequence, such as a control target plasmid, for determining that all of the ingredients are functioning properly. For DNA sequencing, the kit can further comprise sequencing extension primers with homology to one or more sites in the transposable element. Primers can have homology to sequences outside the transposable element (i.e. in a target vehicle).

In the kits, the mutant protein may be added as a purified protein product, may be encoded in the transposable element and produced therefrom, or encoded on vectors separate from the transposable segment, to be produced *in vivo*.

It is to be understood that the invention encompasses transposable systems with varying degrees of reduction of target site specificity from the wild-type which are useful for the purposes of the invention described herein.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1. Papillation phenotypes of the TnsC gain-of-function mutants. Cells were patched on MacConkey lactose plates and photographed after three days' incubation at 30°C. TnsA+B was present in each strain; the TnsC species present is indicated below each patch.

Figure 2. Amino acid changes in the TnsC mutants. The TnsC protein sequence (SEQ ID NOS.1 and 2) is cartooned, with the residues altered in the Class I mutants indicated above the protein and the Class II mutants below the protein. Hatched boxes represent Walker A and Walker B motifs.

Figure 3. TnsC mutants promote transposition to the chromosome. Frequencies of transposition of miniTn7-Km^R from a λ phage to the chromosome were measured by the λ hop assay. TnsA+B was present in each strain; the TnsC species present is indicated below each column.

Figure 4. TnsC mutants promote transposition to conjugable plasmids. Frequencies of transposition of miniTn7-Km^R from the chromosome to the conjugable target plasmid pOX-G were measured by the mating-out assay. TnsA+B was present in each strain; the TnsC species present is indicated below each column.

Figure 5. The substrates, intermediates and products of Tn7 transposition. One substrate is a donor plasmid containing a miniTn7 element which contains the essential cis-acting sequences at each end for transposition. The other substrate is a target plasmid. Transposition initiates with a double strand break at either end of the element, followed by a second break at the other end to generate an excised linear transposon. This excised transposon is then joined to the target DNA to form a simple insertion.

Figure 6. Analysis of Tn7 transposition reactions on a agarose gel. The donor plasmid, a pBR derivative, contained a miniTn7 element containing a kanamycin gene and the target plasmid contained an *attTn7* site. Recombination reactions were carried out as described, the DNAs isolated from the reaction mixture by phenol extraction, digested with a restriction enzyme that cuts once in the donor backbones, displayed by electrophoresis on an agarose gel, transferred to a membrane by electrotransfer and hybridized with a probe specific for the miniTn7 element. Lane 1: TnsA+B; Lane 2: TnsA+B + Cwt; Lane 3: TnsA+B + CE233K; Lane 4: TnsA+B + $CS^{401Y\Delta402}$, Lane 5: Tns(A+B)+$C^{A225V.}$

Figure 7. Tn7 insertion mediated by TnsA+B + C^{A225V} occurs at many different sites in a target DNA. *In vitro* transposition reactions using TnsA+B + C^{A225V} were carried out and the DNAs isolated by phenol extraction and ethanol precipitation. A PCR reaction using the transposition products as a template was then carried out in which one primer (NLC 209) (SEQ ID NO:8) complementary to a sequence on the target DNA and another primer NLC 95 (SEQ ID NO:7) complementary to the left end of Tn7. The length of the PCR products will vary depending on the position of the Tn7 insertion, for example, insertions being closer to the target primer will be short (insert 1) and those more distant will be longer (insert 2). The products of the *in vitro* reaction were then displayed on a denaturing acrylamide gel by electrophoresis, transferred from the gel to membranes and analyzed by hybridization to a radioactively labeled probe that hybridizes to Tn7 sequences on one end of the transposon.

Figure 8. Analysis of distribution of insertions in different regions of the plasmid. Tn7 displays little target site selectivity at many regions of a target. *In vitro* transposition reactions were carried out and the products used as a template for PCR reactions as described above except for the target primer. In these experiments, one primer in the end of Tn7 (NLC 95) (SEQ ID NO:7) was used and in separate reactions primers from several different positions in the target DNA were used.

Figure 9A-C. Structure of Tn7 donor plasmids. A. A plasmid contains a miniTn7 element in which the essential cis-acting sequences at the element termini flank a selectable marker. The translocation of the element can be readily followed by hybridization to a miniTn specific probe. Many different kinds of information could be inside the ends as a selectable (or identifable marker, for example, an antibiotic resistance gene. If the products of transformation are to be recovered *in vivo*, it is convenient to remove unreacted donor DNA by digestion with a restriction enzyme that is selective for the donor backbone; alternatively a conditional replicon can be used. B. Sequence of Donor plasmid pEM delta R.adj to 1 (SEQ ID NO:3). Plasmid carries a 1625 bp mini-Tn7 element: 199 bp of Tn7R and 166 bp of Tn7L flank a Kan gene with *SalI* sites at the junctions. The backbone is pTRC99 (Pharmacia); mini-Tn7 plus flanking host DNA was cloned into the *SmaI* site. C. A commonly used derivative is pEM-Δ, (SEQ ID NO:4) a pBR plasmid containing a kanamycin mTn7 element.

Figure 10A-B. Tn7 target plasmids. A. Sequence of Target plasmid pER 183 (SEQ ID NO:5). This 8.9 kb pACYC184 derivative carries chloramphenicol resistance, a p15A origin of replication, and inserts carrying *mcrB, mcrC, hsdS*, and a segment of phage f1. A large target was used to detect preference of moderate complexity (up to four bp preferences should be detectable). In addition, different segments of the plasmid vary in G+C content from 35% to 68%, so that any preference the transposition system might display for a particular G+C content might be revealed. B. The major targets used in this work are pRM2 (SEQ ID NO:6), a 3190 bp pBR derivative containing at *attTn7* segment and pER183 (SEQ ID NO:5), a pACYC derivative containing several *E. coli* genes.

Figure 11. Diagram of sequencing runs used to ascertain the positions of 63 insertions of mini-Tn7 into pER183 (SEQ ID NO:5). Numbers at the top refer to coordinates on the sequence of pER183 (SEQ ID NO:5) displayed in Fig. 11B. Arrows indicate the direction of primer extension; arrow stems cover the sequence obtained from the run. Arbitrary

numbers attached to the arrows assigned by the sequence assembly program AUTOASSEMBLE.

Figure 12. Graph of the observed distribution of insertions in 100-bp intervals of pER183 (SEQ ID NO:5), and the distribution expected if the distribution were random. On the abcissa is the number of insertions per interval; on the ordinate is the number of intervals that exhibit that number of insertions. Crosses show the expected values for a random (Poisson) distribution of insertions along the sequence; diamonds show the observed values.

Figure 13. The base composition of the 5bp sequences duplicated by the process of Tn7 insertion for the 63 sites examined. On the abcissa, sequence positions are numbered relative to the right end of Tn7 (Tn7R) such that position 1 is immediately adjacent, position 5 is 5bp away (see diagram below the graph). On the ordinate is the number of instances of a particular base at that position. All bases are well represented at all sites.

Figure 14. Effect of four methods of stopping the transposition reaction in preparation for introduction into cells. Results for four replicates (abcissa) of each of four stop methods (z axis), reported as number of transformants per 1/50th of the total reaction (ordinate). Treatments were: no treatment; heat treatment at 65°C for 20 min; heat treatment at 75°C for 10 min; and phenol extraction followed by ethanol precipitation. Heat treatment at 75°C but not 65°C allows effective recovery.

Figure 15. A second experiment displaying the effect of three methods of stopping the transposition reaction in preparation for introduction into cells. Results are shown for two replicates of each of three stop methods (abcissa) for four doses of two different aliquots of TnsB (z axis), reported as number of transformants per 1/25th reaction. Treatments were: heat treatment at 75°C for 10 min; ethanol precipitation alone; or heat treatment at 65°C for 20 min. On the z axis, two aliquots (1- or 2-) of TnsB were used, in four different doses,

1µl 1.5µl 2µl or 3µl. The row labeled 1-2, for example, employed aliquot 1 and used 2µl of it. Heat treatment at 75°C but not 65°C allows effective recovery. This experiment also illustrates the dose-response to TnsB.

Figure 16. Effect of two methods of storing proteins on the efficiency of the transposition reaction. Abcissa displays the storage conditions tested: "individually", TnsA, TnsB and TnsC proteins stored individually in separate tubes at -70°C; "as a mixture (A2a)", TnsA, TnsB and TnsC proteins stored together as a mixture at -70°C. Ordinate displays the number of transformants per 1/50th of the total reaction. Each treatment was tested in quadruplicate.

Figure 17. Effect of three methods of storing proteins on the efficiency of the transposition reaction. Abcissa displays the storage conditions tested: "individually", TnsA, TnsB and TnsC proteins stored individually in separate tubes at -70°C; "as a mixture, –70 (A2a)", TnsA, TnsB, and TnsC proteins stored together as a mixture at -70°C "as a mixture, –20 (A2b)", TnsA, TnsB and TnsC proteins stored together as a mixture at -20°C. Ordinate displays the number of transformants per 1/50th of the total reaction. Each treatment was tested in quadruplicate.

Figures 18A-18B. 18A. nucleotide sequence of TnsC (SEQ ID NO:1). 18B. Amino acid sequence of TnsC (SEQ ID NOS:1 and 2).

DETAILED DESCRIPTION OF THE INVENTION

In the art, the term "transposon" encompasses a segment flanked by particular cis-acting sites that are required for mobilization to occur, together with the genes that specify the proteins that act on those cis-acting sites to mobilize the segment defined by them, whether or not the protein-encoding genes lie between the sites mentioned. For example, according to the present invention, a Tn7 transposon can correspond to the wild type transposon except that the transposon encodes a mutant TnsC. This transposon thus provides the protein products required for mobilization. However, an entire transposon is not necessary to practice the invention. Thus, the term "transposon derivative", "transposable element", or "insertable element" as used herein can also refer to DNA minimally comprising the cis-acting sites at which the trans-acting proteins act to mobilize the segment defined by the sites. It is also understood that the sites may contain intervening DNA.

The phrase "transposable system" as used herein encompasses a transposon containing a mutation in a native ATP-utilizing regulatory protein which, when expressed from the transposon, allows for the non-specific target site selectivity or reduced target site selectivity disclosed herein. The phrase also encompasses modifications in which the relevant proteins are not encoded on the transposable element but nevertheless, acts upon it to achieve the objects of the invention. Thus, the system encompasses compositions in which the mutant protein is added to a transposable element that is derived from a transposon but where the element contains less than the full complement of genes. The only limitation on this element is that it contain the cis-acting sequences upon which the mutant protein acts that allows integration of the element into a target DNA. Thus, the system comprises DNA with cis-acting sites (which may contain heterologous DNA sequences) and the trans-acting proteins that employ those sites to mobilize the segment defined by the sites, regardless of how they are organized in DNA. Accordingly, the proteins may be provided in separate plasmids or in purified form.

The term "transposon-derived" as used herein to refer to the mutant protein, refers to a derivative of a protein normally found on the transposon. However, this need not be the naturally occurring protein but can be the protein produced by recombinant or chemical synthetic methods known to those in the art.

The term "transposable element" encompasses both transposons and derivatives thereof. The only limitation on the derivative is that it is capable of integrating into DNA, containing cis-acting sequences that interact with transacting proteins to effect integration of the element.

The invention provides a transposable system that allows simple integration of a transposable element into a given DNA target efficiently and with a relatively low degree of specificity, preferably random specificity. By "relatively" is intended the degree of specificity compared to the wild-type.

The efficiency of integration can vary depending upon the particular use for which insertion is desired. The mutations described herein increase the efficiency of integration compared to the wild-type frequency. The invention encompasses an efficiency of one simple integration event per every 5-10 kilobases. Preferred levels of integration allow multiple simple insertions in different positions in every gene.

Integration is also effected by the degree of specificity that the mutation confers or allows. Thus, specificity relates to the relationship of a target DNA sequence and the transposable system.

A preferred degree of specificity results in an average insertion in every gene. A practical lower limit would be, on average, one insertion per twenty genes.

For sequencing, greater than or equal to 90% of the insertions screened are at different locations (i.e. 10 insertions hit at least 9 different sites) so that almost every template examined gives new information. This is true in DNAs of a variety of different base compositions since possible target DNAs may vary between 20% and 80% G + C. Another way to describe the possible randomness of the system is to say that of 63 insertions, 62 insertion sites were found (around 98% of insertions are at different locations).

For mutagenesis, non-commercial systems have been widely used that yield as little as 10% of insertions at different sites (i.e. 9 of 10 insertions are at the same site). The present invention improves on this level of randomness.

Furthermore, the types of insertions that are relevant to the discussion of frequency are simple insertions.

The invention provides a transposable system with a mutation that provides for efficient, simple insertion and reduced or random target site specificity.

The term "simple insertion" refers to a single copy integration event of the element introduced into the target by double-strand breakage and rejoining.

Although simple insertions (only one copy of the integrant) are preferred, there may be certain embodiments in which more than one copy does not interfere with the purpose of the application, for example some applications of *in vitro* mutagenesis, or is actually desirable (for example, for multiple copies of a heterologous DNA sequence are to be inserted). Accordingly, the invention is not limited to the case in which the transposable system provides for simple insertion only.

In a preferred embodiment of the invention the mutation is in a transposon-derived ATP-utilizing regulatory protein. One can recognize such a protein by its similarity to the TnsC protein of Tn7, that is by its sequence homology, its possession of a protein sequence motif element similar to an ATP binding site motif in other ATP-dependent proteins, or by reconstitution of an *in vitro* transposition system and demonstration of a requirement for nucleotides in that *in vitro* transposition system.

In a highly preferred embodiment, the mutation is in the TnsC gene (SEQ ID NO:1) encoding the TnsC protein (SEQ ID NOS:1 and 2) of Tn7. This mutation provides a Tn7 transposon that is capable of relatively non-specific insertion into a given DNA segment.

Thus, the invention is directed to insertion of the Tn7 transposable element but is not limited to this transposable element. Accordingly, the invention can be practiced with transposable elements related to Tn7 in that transposition occurs by means of an ATP-mediated process. Thus, mutations in the ATP-utilizing proteins in such transposons is

contemplated in this disclosure. Accordingly, transposons with ATP-utilizing regulatory proteins in addition to Tn7 are encompassed in the invention. Examples of such transposons are Tn5090/Tn420; the transposon-encoded transposition proteins are TniA, TniB, and TniQ. The TniB would be the ATP-utilizing protein.

Another class of transposon is encompassed by the invention in which it is possible to increase the frequency by altering the ATP-utilizing proteins examples are Tn552 and IS21.

The invention provides for the insertion of the transposable elements described herein into any DNA segment of any organism. Moreover, the invention also provides for the insertion into any synthetic DNA segment.

Insertion of the transposable element can be *in vivo*. In this case, the transposable element is introduced into a desired host cell, where it inserts directly into DNA in that cell. The only limitation is that the transposable element be capable of insertion in the specific host cell DNA. Thus, as long as the proteins required for transposition can be expressed in a desired cell, this cell can provide a host for insertion of the transposable element into any DNA found in that host cell.

Insertion can be for the purpose of gene inactivation. Gene inactivation is useful for genetic analysis (e.g. gene function).

Genetic analysis includes:

assessment of the phenotype of a null allele (not expressing functional protein due to interruption of the gene by the transposable segment); assessment of the consequences of insertion of particular active DNA structures or sequences for genetic properties of chromosomes or their parts, such as but not limited to accessibility to Dnase I or to footprinting reagents, or expression or silencing of nearby transcribable genes, or for activity of genetic or epigenetic processes such as, but not limited to homologous recombination, chemical mutagenesis, oxidative DNA damages, DNA methylation, insertion of proviruses or retroposons; assessment of protein domain structure via creation of multiple interruption points within a gene for a multidomain protein, wherein a gene product missing one or more domains of the multidomain protein might exhibit partial

activity or activities, including antigenic activities or immunodominant epitopes [randomness is paramount here, many insertion positions are needed if borders are to be defined accurately]; assessment of expression pattern via creation of transcriptional fusions of a promoter in the target to a reporter (e.g. beta galactosidase or green flourescent protein or chloramphenicol transacetylase or luciferase) within the transposable segment; assessment of expression pattern via creation of translational fusions of a portion of a gene product encoded by a target to a gene product or an antigenic peptide encoded by the transposable segment (e.g. beta galactosidase or an epitope tag or an affinity tag); assessment of operon structure, in which interruption of transcription by insertion upstream of a gene results in altered expression of a gene without disrupting the coding sequence of that gene; gratuitous expression of a gene, in which transcription from a promoter within the transposable segment results in expression of a gene downstream of the position of insertion of the transposable segment, with or without regulation of transcription of the promoter within the transposable segment; gratuitous expression of a protein fusion, in which transcription from a promoter within the transposable segment results in translation of a protein beginning within the transposable segment and proceeding toward the outside of the transposon, then continuing into the gene within which the transposable segment is inserted, resulting in a fusion of the transposon-encoded protein with the target protein; assessment of the consequences of introducing into the host cell any transcript or gene product entirely encoded within the transposable segment, especially where it is desirable to assess position-effects (the consequences not only of expression but of expression in different positions within the genome).

Insertion can also be for the purpose of introducing heterologous DNA sequences into the DNA of a host cell. The DNA in the host cell in which the insertion occurs can be the host genomic DNA or extrachromosomal elements. This includes both naturally-occurring elements and elements introduced exogenously.

Heterologous genes that can be introduced *via* the insertion include reporter genes. DNA sequences can also be introduced that provide physical markers in a chromosome. Insertion can also be used as a simple way to recover the host DNA that is flanking the

inserted element. Genomic DNA is cut with restriction enzymes and the insertion plus the flanking DNA is then cloned.

Another utility or another application of the invention is to analyze the interaction of various non-transposition proteins with a DNA sequence, for example, DNase footprinting of repressors bound to DNA. A further use is to study the structure of genomic chromatin i.e., the state at which DNA is actually found in the cell.

A further advantage in using Tn7 and similar transposons is that of double end or "concerted" joining. Accordingly, Tn7 inserts in a "cut and paste" manner with both ends of the transposon being joined to the target DNA.

Insertion can also be *in vitro*. *In vitro* insertion provides an advantage over insertion *in vivo*. Using *in vitro* insertion, the transposable element can be placed in any DNA target and that target then introduced into a host cell where it can integrate or replicate. Accordingly, this greatly expands the host cell range.

Targets for insertion, accordingly, include DNA fragments, plasmids and other extrachromosomal elements capable of replication in prokaryotic and/or eukaryotic host cells. Given the array of plasmids available, potentially any cell can be used as a host for an insertion target containing a transposable element that was introduced into the target *in vitro*. The target can be based on a bacterial plasmid, bacteriophage, plant virus, retrovirus, DNA virus, autonomously replicating extra chromosomal DNA element, linear plasmid, mitochondrial or other organelle DNA, chromosomal DNA, and the like.

When introduced into the host cell, the target can be maintained as an autonomously replicating sequence or extrachromosomal element or can be integrated into host DNA. When integrated, integration can occur by homologous recombination or by means of specific integration sequences such as those derived from retroviruses, DNA viruses, and the like.

It may be, but is not necessarily, desirable to obtain replication of the target in the host cell. A specific application in which this is desirable is the case in which a transposable element is used as a component for introducing primer binding sites for DNA sequencing.

Accordingly, in a highly preferred embodiment of the invention, a transposable element is introduced into a target containing a DNA segment for which a sequence is desired. This target is then introduced into a host cell where it is allowed to replicate, thus producing sufficient copies to allow DNA sequencing using a primer specifically recognizing a sequence in the target.

In one embodiment of this method, the primers recognize one or both ends of the transposable element such that sequencing can proceed bidirectionally from the transposable element insertion site into the surrounding DNA. The target may be composed entirely of DNA segments for which the sequence is required or may simply contain subsequences for which a sequence is required. In this aspect the only limitation on the target is that it is able to replicate in the host cell (and therefore contains sequences that allow this to occur).

It is also highly desirable that the target have a selection marker in order to eliminate the background in host cells containing the target without the insertion of the transposon.

An alternative way to eliminate this background, however, is to provide a method for disabling a target that has not received an insertion so that it is unable to replicate in the host cell and is thus diluted out during host cell culture. Accordingly, the transposable element itself could contain an origin of replication for the host cell. Thus targets not receiving an insertion would be unable to replicate. An insertion could also result in the formation of functional replication sequences. The target could also contain a heterologous conditional origin, such as the R6K origin, that cannot replicate without the pir protein. The person of ordinary skill in these arts would be aware of the various methods for constructing targets with the (in)ability to replicate in a specific host cell.

It is also possible, however, to use the transposable elements described herein for DNA sequencing without the *in vitro* insertion described above. Insertion could be directly accomplished in host cell DNA and then the DNA containing the insertion removed from the host. This DNA segment could then be replicated although it does not necessarily have

to be if the host has produced sufficient copies for sequencing. Accordingly, sufficient numbers of the segment with the insertion sequence could then be sequenced as above.

An example of the case in which the DNA segment receiving an *in vivo* insertion would not need to be further replicated in another host is, for example, a case in which the insertion occurs in a sequence capable of being amplified directly in the host cell. This could be a plasmid containing an amplifiable marker, such as the *dhfr* gene, the cell being grown in a selective medium containing methotrexate. The person of ordinary skill in the art would know the various methods for amplifying DNA segments using selectable markers. The selectable marker could be introduced on the transposon but would not necessarily need to be.

In a further DNA sequencing protocol, the primers that are used facilitate DNA segment amplification by the PCR reaction. For example, a primer can be used that recognizes an end of the transposable element with the second primer being found in the target DNA sequence. The primer could be based on random sequences or on known sequences deliberately placed in the target vehicle. Thus the target vehicle could contain a characterized plasmid (as an example) in which the sequences are known. In this instance, primers can be designed to hybridize to any area within the plasmid, the segment to be sequenced being between the transposon and the second primer site in the target vehicle.

In accordance with the above-described embodiment, the invention is also directed to kits for performing transposable element insertion *in vitro*. As described, such insertions can be used to provide priming sites for DNA sequence determination or to provide mutations suitable for genetic analysis or both.

Essential components in the kit are gene products allowing transposition that are normally encoded on the transposable element or their functional equivalents. A further component is a transposable element donor vehicle. This nucleic acid vehicle provides the transposable element to be inserted into a given specific target. The transposable element donor is preferably DNA but could encompass RNA, being operable via a cDNA copy. Preferred DNA vehicles include, but are not limited to, bacterial plasmids. Other vehicles include any DNA that can be isolated in super coiled form or placed into a super coiled

configuration by the use of topoisomerases, for example, bacteriophaged DNA, autonomously replicating molecules from eukaryotes or archae, or synthetic DNA that can be ligated to form a topologically closed circle.

Optional components of the kit include one or more of the following: (1) buffer constituents, (2) control target plasmid, (3) sequencing primers. The buffer can include any buffer suitable for allowing the transposition activity to occur *in vitro*. A preferred embodiment is HEPES buffer. A specific disclosed embodiment is included in the exemplary material herein.

Preferred donor plasmids do not need to be destroyed before introducing transposition products into commonly used bacterial and preferably *E. coli* strains. These vectors do not replicate without regulatory genes not provided by the host cell which allow a functional replication origin. An example is the pir gene which is present only in specially constructed strains, having been derived from the plasmid R6K. In this way, artifactual background consisting of cells transformed with both the donor DNA and the target DNA without any transposition having occurred is eliminated. As discussed herein, there are other ways to do this such as restriction digestion of the donor DNA but not of the target or transposable segment or deletion and titration of the transposition reaction so that there are more cells than DNA molecules in the transformation step. However, these are not preferred.

The control target plasmid does not contain the transposable element and does contain transposable element integration site. The purpose is to assure that the reaction is not inhibited by a contaminant in non-kit ingredients (introduced by the kit user); i.e. it ensures that all components allow optimal insertion.

Sequencing primers include, but are not limited to primers that have homology with both ends of the transposable element and, as such, allow sequencing to proceed bidirectionally from the ends of the transposable element. However, primers could be made to any area within the transposable element or within the target vehicle itself as long as extension is allowed into the DNA segment to be sequenced. Kits designed for allowing

sequencing by the PCR reaction may also include a second primer that allows the amplification of the sequence between the first and second primers.

The control target plasmid preferably contains a selectable marker for recovery of the desired DNA segment from a specific host cell. It is understood that, when using the kit, the target DNA does not carry the same selectable marker as the control target nucleic acid.

A fourth optional component of a kit is target DNA itself. Target DNA that might be desirable would include but is not limited to purified chromosomal DNA, total cDNA, cDNA fractionated according to tissue or expression state (e.g. after heat shock or after cytokine treatment other other treatment) or expression time (after any such treatment) or developmental stage, or plasmid, cosmid, BAC, YAC or phage library of any of the foregoing DNA samples, especially such target DNA from important study organisms such as *Homo sapiens, Mus domesticus, Mus spretus, Canis domesticus, Bos, Caenorhabditis elegans, Plasmodium falciparum, Plasmodium vivax, Onchocerca volvulus, Brugia malayi, Dirofilaria immitis, Leishmania, Zea maize, Arabidopsis thaliana, Glycine max, Drosophila melanogaster, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Neurospora, Escherichia coli, Salmonella typhimurium, Bacillus subtilis, Neisseria gonorrhoeae, Staphylococcus aureus, Streptococcus pneumonia, Mycobacterium tuberculosis, Aquifex, Thermus aquaticus, Pyrococcus furiosus, Thermus littoralis, Methanobacterium thermoautotrophicum, Sulfolobus caldoaceticus*, and others.

Other suitable selectable markers include chloramphenicol resistance, tetracycline resistance, spectinomycin resistance, streptomycin resistance, erythromycin resistance, rifampicin resistance, bleomycin resistance, thermally adapted kanamycin resistance, gentamycin resistance, hygromycin resistance, trimethoprim resistance, dihydrofolate reductase (DHFR), GPT; the URA3, HIS4, LEU2, and TRP1 genes of S. cerevisiae.

There may be certain instances in which it is desired to introduce primer binding sites other than those naturally found in the transposable element or in the insertion vehicle. In this case, the transposable element can be used as a vehicle for introducing any desired primer or primers. An example of when the use of exogenous primers may be desirable is

the case in which the transposable element ends form a secondary structure that interferes with sequencing, or cases in which there is a similarity of sequence between the two ends of the transposable element, and cases in which the only practical binding sites in the transposable element are so far internal that they undesirably curtail the amount of nucleotides that can be sequenced from that site.

The invention also generally encompasses compositions containing an ATP-dependent DNA binding protein encoded by a transposon, the protein containing a mutation conferring reduced target site specificity, preferably random target site insertion.

The protein is isolated from a biological preparation produced *in vivo* or *in vitro*. Thus, the protein is purified or substantially purified from cellular components with which it is found *in vivo*. When produced *in vitro*, the protein may also be purified or substantially purified from the other components used to produce it.

In preferred embodiments the protein is the TnsC protein (SEQ ID NOS:1 and 2).

In a specific disclosed embodiment, the protein contains a valine at amino acid number 225.

The invention is also directed to compositions containing the protein described herein and the transposable element substrate on which the protein acts to cause insertion.

Compositions can also include target DNA into which the transposable element is capable of being inserted.

The mutant proteins of the present invention include the naturally occurring proteins encoded by a transposon as well as any substantially homologous and/or functionally equivalent variants thereof. By "variant" protein is intended a protein derived from the native protein by deletion (so-called truncation) or addition of one or more amino acids to the N-terminal and/or C-terminal end of the native protein; deletion or addition of one or more amino acids at one or more sites in the native protein; or substitution of one or more amino acids at one or more sites in the native protein. Such variants may result from, for example, genetic polymorphism or from human manipulation. Methods for such manipulations are generally known in the art.

For example, amino acid sequence variants of the polypeptide can be prepared by mutations in the cloned DNA sequence encoding the native protein of interest. Methods for mutagenesis and nucleotide sequence alterations are well known in the art. See, for example, (37) (38) (39) (40); U.S. Patent No. 4,873,192; and the references cited therein; herein incorporated by reference. Guidance as to appropriate amino acid substitutions that do not affect biological activity of the protein of interest may be found in the model of Dayhoff *et al.* (1978) (41) in *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, D.C.), herein incorporated by reference. Conservative substitutions, such as exchanging one amino acid with another having similar properties, may be preferred.

In constructing variants of the protein of interest, modifications to the nucleotide sequences encoding the variants will be made such that variants continue to possess the desired activity. Obviously, any mutations made in the DNA encoding the variant protein must not place the sequence out of reading frame and preferably will not create complementary regions that could produce secondary mRNA structure. See EP Patent Application Publication No. 75,444.

Thus nucleotide sequences of the invention and the proteins encoded thereby include the naturally occurring forms as well as variants thereof. The variant proteins will be substantially homologous and functionally equivalent to the native protein. A variant of a native protein can be "substantially homologous" to the native protein when at least about 80%, more preferably at least about 90%, and most preferably at least about 95% of its amino acid sequence is identical to the amino acid sequence of the native protein. However, substantial homology includes high homology in the catalytic or other conserved functional regions with possible low homology outside these. By "functionally equivalent" is intended that the sequence of the variant defines a chain that produces a protein having substantially the same biological effect as the native protein of interest. Thus, for purposes of the present invention, a functionally equivalent variant will confer the phenotype of activating transposition with reduced target site specificity, preferably random. Such

functionally equivalent variants that comprise substantial sequence variations are also encompassed by the invention.

The invention also encompasses compositions containing a transposable element containing DNA sequence encoding an ATP-utilizing regulatory protein, the protein containing a mutation that confers reduced target site specificity and preferably random insertion.

In preferred embodiments of the invention, the transposable element is a Tn7 transposable element.

In specific disclosed embodiments, the mutation is valine as amino acid number 225 in the TnsC protein.

The invention also encompasses compositions containing the above-described transposable element and a given DNA segment intended to be the target for insertion of the transposable element.

The invention, accordingly, is directed to DNA into which has been inserted the transposable element containing the mutation described herein that confers simple, efficient insertion with reduced target site specificity or random target site insertion. The DNA in this composition, in one embodiment, is capable of being introduced into a cell in which it can exist as an extrachromosomal element or as an integration element into cellular DNA.

The invention is also directed to DNA segments encoding the mutant proteins disclosed herein, vectors containing these segments and host cells containing the vectors. The vectors containing the DNA segments may be used to propagate (i.e. amplify) the segment in an appropriate host cell and/or to allow expression from the segment (i.e. an expression vector). The person of ordinary skill in the art would be aware of the various vectors available for propagation and expression of a cloned DNA sequence. In a preferred embodiment, a DNA segment encoding mutant TnsC protein is contained in a plasmid vector that allows expression of the protein and subsequent isolation and purification of the protein produced by the recombinant vector. Accordingly, the proteins disclosed herein can be purified following expression from the native transposon, obtained by chemical synthesis, or obtained by recombinant methods.

Relevant compositions, accordingly, include expression vectors for the mutant protein alone or in combination with expression vectors for the other proteins necessary for insertion of a transposable element. Such compositions may further comprise the transposable element to be acted upon by the proteins. Such mixtures are useful for achieving *in vivo* insertion, among other things.

The invention further encompasses kits containing the above-described compositions.

Tn7 can be obtained as strain ATCC 29181; a K-12 derivative carrying the resistance transfer factor R483; originally identified as carrying a transposon in Barth *et al. J. Bacteriol. 125*:800-810 (1976). The sequence of Tn7 is Genbank entry ISTN7TNS, Assession no. X17693; reported in Flores *et al. Nucleic Acids Res. 18*:901-11 (1990).

Having now generally described this invention, the same will be further described by reference to certain specific examples which are provided herein for purposes of illustration only and are not intended to be limiting unless otherwise specified.

EXPERIMENTAL

Example 1

Materials and Methods

Media, chemicals, and enzymes: LB broth and agar were prepared as described (42). Trimethoprim selection was on Isosensitest agar (Oxoid). Lac phenotypes were evaluated on MacConkey lactose agar (Difco). Antibiotic concentrations used were 100 μg/ml carbenicillin (Cb), 30 μg/ml chloramphenicol (Cm), 7.5 μg/ml gentamycin (Gn), 50 μg/ml kanamycin (Km), 10 μg/ml nalidixic acid (Nal), 20 μg/ml tetracycline (Tet) and 100 μg/ml trimethoprim (Tp). Hydroxylamine was purchased from Sigma. DNA modifying enzymes were purchased from commercial sources and used as recommended by the manufacturer.

Bacterial strains, phages and plasmids: BR293 is *E. coli* F⁻ Δ(lac-pro) *thi* rpsL Δ(*gal* - λ*G*) + lacZ pL cI+$_{434}$ pRS$_7$ (43) (44). BR293 is identical to NK8027 (45), and was provided by Nancy Kleckner. NLC51 is *E. coli* F⁻ araD139 Δ(argF-lac) U169 rpsL150 relA1 flbB5301 deoC1 ptsF25 rbsR Val^R recA56 (46). CW51 is *E. coli* F⁻ ara arg Δlac-proXIII recA56 Nal^R Rif^R (11). λKK1 is lambda 780 hisG9424::Tn10 del16 del17::*attTn7*::miniTn7-Km^R (47). Tns transposition proteins were provided by pCW15 (tnsABC), pCW23 (tnsD), pCW30 (tnsE), or pCW4 (tnsABCDE) (11). Target plasmids were derivatives of pOX-G, a conjugable derivative of the F plasmid that carries Gn^R (48). pOX-*attTn7* carries a (-342 to +165) *attTn7* sequence (16). The immune plasmid pOX-*attTn7* EP-1::miniTn7-Cm^R was made by transposing miniTn7-Cm^R (47) onto pOX-*attTn7* using TnsABC+E to direct the insertion into a non-*attTn7* position. Construction of the immune target plasmid pOX-G::miniTn7-*dhfr* is described below. The transposon donor plasmid for the papillation assay was pOX-G::miniTn7lac, containing promoterless lacZY between the transposon ends (50). The high copy transposon donor for mating-out assays was pEMΔ, containing miniTn7-Km^R (23).

Manipulation and characterization of DNA: Phage and plasmid isolation, transformation, and standard cloning techniques were performed as described in (40). Conjugation and P1 transduction were performed as described in (42). DNA sequencing was done on an automated ABI sequencer. Two plasmids were constructed in this work: (1) pOX-G::miniTn7-*dhfr*. MiniTn7-*dhfr* was constructed by replacing the Km^R cassette in pLA1 (16) with a *dhfr* cassette from pSD511 (28), which had been amplified by PCR to add flanking *SalI* sites. The PCR fragment was ligated into the TA vector (Invitrogen), the *dhfr* cassette was then removed by SalI digestion and inserted into the SalI site of pLA1, replacing the Km^R gene. The resulting plasmid was transformed into NLC51+pCW4+pOX-G, and grown for several days to allow transposition to occur. pOX-G plasmids which had received a miniTn7-*dhfr* insertion were identified by mating into CW51 and selecting for Tp^R.

Mutagenesis of tnsC: The TnsABC plasmid pCW15 was exposed to 1M hydroxylamine hydrochloride in 0.45 M NaOH (final pH approximately 7.0) at 37°C for 20 hours (ROSE *et al.* 1990). The DNA was recovered by multiple ethanol precipitations, and *PvuII-SphI* fragments containing mutagenized tnsC were subcloned into untreated pCW15, replacing the wild-type TnsC (SEQ ID NO.1). These plasmids were then introduced into CW51 + pOX-G::miniTn7lac by electroporation, and transformants were selected on MacConkey lactose plates containing Cm. The plates were incubated at 30°C for 3-4 days, and screened for the emergence of Lac^+ papillae, indicating transposition of miniTn7lac.

λ hop transposition assay: Tn7 transposition was evaluated in NLC51 strains into which tns functions were introduced by transformation, and pOX-G was introduced by conjugation (for Figure 5). The protocol of (47) was followed: Cells were grown in LB and 0.2% maltose at 37°C to an OD_{600} of 0.4-0.6 and then concentrated to 1.6×10^9 cells/ml by centrifugation and resuspension in 10 mM $MgSO_4$. 0.1 ml cells were combined with 0.1 ml λKK1 containing miniTn7-Km^R at a multiplicity of infection of 0.1 phage per cell. The infection proceeded for 15 min at 37°C, and was terminated by the addition of 10 mM sodium citrate in 0.8 ml LB. Cells were allowed to recover with aeration for 60 minutes at

37°C, and then spread on plates containing Km and citrate. Transposition frequency is expressed as the number of Km^R colonies/pfu of λKK1.

Mating-out transposition assay: Tn7 transposition was evaluated in the derivatives of BR293 used to monitor SOS induction (Table 3), or in NLC51 strains into which tns functions were introduced by transformation, and pOX-G or pOX-G::miniTn7-*dhfr* were introduced by conjugation. MiniTn7-Km^R was present in the NLC51 strains either in the chromosomal *attTn7* site (Figure 4 and Table 1) or the high copy plasmid pEMΔ (Table 2) (SEQ ID NO:4). The protocol was adapted from (11): The donor strains described above and the recipient strain CW51 were grown at 37° to an OD_{600} of 0.4-0.6 with gentle aeration. Donors and recipients were mixed at a ratio of 1:5, and growth was continued for another hour. Mating was disrupted by vigorous vortexing, and the cells were diluted and plated. The total number of exconjugants was determined by selection on GnNal plates. Tn7-containing exconjugants were selected on TpNalplates, and miniTn7-Km^R exconjugants were selected on KmNal plates. Transposition frequencies are expressed as the number of Tp^R- or Km^R-exconjugants/total number of exconjugants.

Results

Isolation of the TnsC gain-of-function mutants: To focus on the relationship of TnsC and the target DNA, the inventor isolated gain-of-function TnsC mutants that activated the TnsA+B transposase in the absence of TnsD or TnsE. Since overexpression of wild-type TnsC does not relieve the requirement for TnsD or TnsE (11), these gain-of-function mutations were predicted to affect the biochemical properties of TnsC, rather than its expression or stability.

A visual assay for Tn7 transposition (50) (51) was used to identify mutants. This assay uses a miniTn7lac element which carries promoterless lacZY genes between the cis-acting sequences at the transposon ends. The miniTn7lac element is located in a transciptionally silent position on a donor plasmid; cells containing this plasmid are phenotypically Lac^-. When Tns functions are provided *in trans*, miniTn7lac can transpose

to new sites in the *E. coli* chromosome. Some of those transposition events place the element downstream from active promoters, resulting in increased lacZ expression. This is observed on MacConkey lactose color indicator plates as the emergence of red (Lac^+) papillae in an otherwise white (Lac^-) colony. Therefore, the number of papillae reflects the amount of transposition which occurred during the growth of that colony.

Cells containing miniTn7lac and various Tns functions were patched on color indicator plates (Figure 1). Virtually no Lac^+ papillae were seen in cells containing only $TnsABC^{wt}$. Cells containing $TnsABC^{wt}$+E produced many Lac^+ papillae. Southern blotting demonstrated that $TnsABC^{wt}$+E papillae result from translocations of miniTn7lac to a variety of chromosomal locations rather than from intramolecular rearrangements of the donor plasmid (50). Most $TnsABC^{wt}$+D events are silent because there is no appropriately oriented promoter adjacent to *attTn7* (50) (52).

This visual assay was used to screen for TnsC mutants that had acquired the ability to activate Tn7 transposition in the absence of TnsD or TnsE. Randomly mutagenized tnsC was cloned into a plasmid containing tnsAB. These tns genes were introduced into cells containing miniTn7lac. Six gain-of-function TnsC mutants were identified (Figure 1).

Transposition activated by these TnsC mutants still required the TnsA+B transposase and intact transposon ends. The papillation phenotypes of the TnsC mutants varied considerably, suggesting that different mutants were activating different amounts of miniTn7lac transposition. Several TnsC mutants promoted more transposition than $TnsABC^{wt}$+E. $TnsABC^{S401Y\Delta402}$ achieved the highest level of transposition.

The amino acid changes responsible for the mutant phenotypes were determined by DNA subcloning and sequencing. tnsC encodes a protein of 555 amino acids, with Walker A and B motifs in the amino-terminal half of the protein (53). Walker A and B motifs have been implicated by structural and mutational analyses to be directly involved in nucleotide binding and/or hydrolysis in a variety of ATPases and GTPases (37) (55).

The tnsC mutations primarily result in single amino acid substitutions whose locations are scattered across the TnsC protein sequence (Figure 2). TnsC mutants

segregate into two phenotypic classes. Transposition reactions activated by Class I mutants are sensitive to immune targets and the target selection factors TnsD and TnsE. Transposition reactions activated by the Class II mutants are impaired in their responses to these signals. The residues affected in two of the mutants ($TnsC^{A225V}$ and $TnsC^{E233K}$) lie in or very close to the Walker B motif.

TnsC mutants promote intermolecular transposition: The papillation assay is a powerful screen for transposition activity, but it does not necessarily report intermolecular transposition events. Internal rearrangements of the miniTn7lac donor plasmid, which fortuitously place the miniTn7lac element downstream from a promoter, would also produce Lac^+ papillae. Therefore, the inventor investigated whether the TnsC mutants facilitate the TnsA+B transposase to do intramolecular recombination, or whether the mutants promote intermolecular transposition.

The λ hop assay measures the translocation of a miniTn7-Km^R element from a replication- and integration-defective λ phage to the bacterial chromosome during a transient infection. The miniTn7-Km^R element carries a kanamycin resistance cassette with a constitutive promoter. Therefore, the λ hop assay reports the total number of transposition events occurring into the chromosome. $TnsABC^{wt}$ had no detectable transposition activity in the λ hop assay. $TnsABC^{wt}$+E generated 2.2×10^{-7} Km^R colonies/pfu (Figure 3). Transposition promoted by $TnsABC^{wt}$+D generated 1.8×10^{-4} Km^R colonies/pfu. All of the TnsC mutants could promote the translocation of miniTn7 - Km^R. $TnsABC^{A225V}$ and $TnsABC^{S404Y\Delta402}$ promoted 8- and 50-fold more transposition than $TnsABC^{wt}$+E. Other TnsC mutants promoted transposition, although not at such levels.

The mating-out assay was used to explore the ability of the TnsC mutants to promote translocations into a different type of target molecule. This assay measures the frequency of transposition of miniTn7-Km^R from the chromosome to pOX-G, a conjugable derivative of the *E. coli* F factor. The $TnsABC^{wt}$+E machinery preferentially selects

conjugable plasmids as targets for transposition, whereas the TnsABC^{wt}+D machinery does not recognize pOX-G unless it contains *attTn7* sequences (ROGERS *et al.* 1986, WADDELL and CRAIG 1988, WOLKOW *et al.* 1996). The TnsC mutants could promote transposition to pOX-G (Figure 4). Thus, the results demonstrate that the gain-of-function TnsC mutants can promote intermolecular transposition.

Effects of the target selection factors TnsD and TnsE: Frequencies of transposition of miniTn7-Km^R from a lambda phage to the chromosome and/or pOX-G were measured in strains containing TnsA+B and the TnsC mutants, either alone or in combination with TnsD or TnsE. The preferred target for TnsE reactions, pOX-G, was introduced by conjugation into strains containing TnsC mutants or the TnsC mutants + TnsE. The distribution of miniTn7-Km^R insertions between the chromosome and pOX-G was determined by mating the pOX-G plasmids from the Km^R products of a lambda hop assay into the Km^s strain CW51, and testing whether Km resistance was plasmid linked. No transposition was detected in strains containing TnsABC^{wt} alone or TnsABC^{E233K} + TnsD.

Response to the target selectors TnsD and TnsE: TnsD and TnsE are required to activate the TnsABC^{wt} machinery and to direct transposition into particular target DNAs (10) (11) (13) (14). The TnsABC^{mutant} machineries, by definition, do not require the inputs of TnsD or TnsE. However, the inventor investigated whether TnsD or TnsE could influence the frequencies or distribution of transposition events promoted by the TnsC mutants.

The λ hop assay was used to evaluate the effects of TnsD and an available *attTn7* site on transposition promoted by the TnsC mutants. All of the mutant reactions were responsive to TnsD+*attTn7*, but those responses varied widely. Reactions activated by TnsABC^{A225V} and TnsABC^{E273K} were strongly stimulated by TnsD+*attTn7*, promoting 500- and 5000-fold more transposition, respectively, in the presence of TnsD+*attTn7* than with TnsABC^{A225V} or TnsABC^{E273K} alone. The remaining mutant reactions were less profoundly influenced by TnsD: TnsABC^{S401F} reactions showed a moderate (50-fold)

stimulation. Reactions activated by $TnsC^{E233K}$, $TnsC^{S401Y\Delta402}$ and $TnsC^{A282T}$ were somewhat inhibited in the presence of TnsD.

The effects of TnsE was also studied by the λ hop assay. In the absence of TnsE, the vast majority of the $TnsABC^{mutant}$ transposition events were targeted to the chromosome. In the presence of TnsE, preferential insertion into pOX-G was observed with some of the TnsC mutants.

These differential responses suggest that the six TnsC mutants are not activating Tn*7* transposition through a single mechanism. Instead, the mutants can be segregated into two classes, based on their ability to respond to TnsD and TnsE. Transposition activated by the Class I mutants -- $TnsC^{A225V}$ and $TnsC^{E273K}$ -- can by stimulated by TnsD and targeted to pOX-G by TnsE. Transposition activated by the Class II mutants -- $TnsC^{E233K}$, $TnsC^{A282T}$, $TnsC^{S401Y\Delta402}$ and $TnsC^{S401F}$ -- is not responsive to the positive effects of TnsD or TnsE or both. By these criteria, $TnsC^{S401F}$ is proposed to be a member of Class II: although $TnsC^{S401F}$-activated reactions are somewhat stimulated by TnsD, the distribution of insertions in $TnsC^{S401F}$-activated reactions is not affected by TnsE. The grouping of the TnsC mutants into these two classes is supported by the differential responses of the $TnsABC^{mutant}$ reactions to immune targets, as described below.

Discussion

Proteins involved in target evaluation: How is an appropriate target for Tn*7* transposition identified? The inventor has hypothesized that TnsC may serve as a "connector" or "matchmaker", linking the transposase and the target DNA in a manner regulated by the ATP state of TnsC (23) (27). TnsC has the biochemical properties necessary for that connection: it can directly interact with target DNA (24) and with the TnsA+B transposase (A. STELLWAGEN and N. L. CRAIG, unpublished results). However, wild-type TnsC (SEQ ID NO:1) is not sufficient to activate transposition.

Instead, Tn7 transposition is dependent on TnsD or TnsE to activate the TnsABC^{wt} machinery and select a target site. TnsD is an *attTn7* binding protein (23) which recruits TnsC to this target. The resulting TnsC-TnsD-*attTn7* complex can then attract the transposase *in vitro* (23). The mechanism by which TnsE activates transposition is not yet known. TnsE might be preferentially localized to conjugating plasmids and subsequently recruit TnsC to those molecules, or TnsE might modify TnsC so that TnsC's binding activity is now directed to those targets. Alternatively, TnsE might modify the transposase directly, without proceeding through TnsC. The results suggest that TnsD and TnsE provide alternative inputs into TnsC, which in turn recruits the TnsA+B transposase to the target DNA.

The successful isolation of TnsC gain-of-function mutants reveals that the TnsABC machinery is capable of engaging target DNA and promoting insertions without TnsD or TnsE. However, the mutant reactions have not mimicked the abilities of TnsD or TnsE to direct transposition into particular targets: transposition activated by the TnsC mutants does not show the preferential insertion into conjugable plasmids seen with TnsE-activated reactions, nor the *attTn7* specificity of TnsD-activated reactions. Therefore, TnsD and TnsE are essential to recognize these positive target signals.

TnsC appears to receive a variety of inputs -- from TnsD, TnsE and from immune targets -- which control its activity. The activity of TnsC can also be influenced by mutation. Six gain-of-function TnsC point mutants have been described in this work, which segregate into two classes. The fact that different classes of TnsC mutants with different transposition activities were recovered is consistent with the hypothesis that there are multiple routes to activating TnsC. The Class I mutants, TnsC^{A225V} and TnsC^{E273K}, enable the TnsABC machinery to execute transposition without sacrificing its ability to respond to both positive and negative target signals. Both are substantial gain-of-function mutants, with TnsABC^{A225V} promoting eight-fold more transposition to the chromosomes than TnsABC^{wt}+E (Figure 3). Transposition activated by these Class I mutants can be profoundly stimulated by TnsD+*attTn7*, or directed to conjugable plasmids by TnsE, as

well as being able to discriminate between immune and non-immune targets. Thus, the gain-of-function phenotypes seen with the Class I mutants have been achieved while preserving the ability of these TnsCs to transduce information between the target DNA and the transposase.

The TnsC mutants which fall into the second class behave much more like constitutively activated versions of TnsC. Some of these mutants also promote considerable amounts of transposition: TnsABC$^{S401Y\Delta402}$ results in 50-fold more transposition to the chromosomes than TnsABC^{wt}+E (Figure 3). However, the nature of the transposition reactions promoted by the Class II TnsC mutants is quite different than those seen with the Class I mutants. Immune and non-immune targets are used essentially equivalently in reactions with the Class II mutants, and TnsD and TnsE are not able to profoundly influence the frequency or distribution of these transposition events. A similar loss of responsiveness to target signals is seen when Tn7 transposition is activated by nonhydrolyzable ATP analogs *in vitro*. Transposition can still occur when TnsC's ATPase activity is blocked with AMP-PNP, but those transposition events no longer require TnsD and are no longer targeted to *attTn7* (BAINTON *et al.* 1993). Instead, any DNA molecule, including immune targets, can serve as a target for Tn7 insertion. Thus, TnsABC transposition can be constitutively activated by AMP-PNP or by the Class II TnsC mutants. It is noteworthy that the amino acid affected in TnsC^{E233K} lies in one of TnsC's ATP motifs.

Comparison to other elements: The use of an ATP-dependent protein such as TnsC to regulate target site selection is not unique to Tn7. Bacteriophage Mu transposition is also profoundly influenced by its ATP-utilizing protein MuB. MuB is an ATP-dependent DNA binding protein (57) (MAXWELL *et al.* 1987) which is required for efficient transposition *in vivo* (58) (59). *In vitro*, the MuA transposase preferentially directs insertions into targets that are bound by MuB (60) (61) (19). Although there is no particular sequence specificity to MuB binding, its distribution on DNA is not random: MuB binding to target molecules that already contain Mu sequences is specifically

destabilized through an ATP-dependent mechanism (19). Therefore Mu, like Tn7, recognizes and avoids immune targets; moreover, MuB and $TnsC^{A225V}$ appear to play functionally similar roles in regulating transposition.

Mu and Tn7 belong to a family of transposons which encode proteins with ATP binding/hydrolysis motifs; other members include IS21 (35) (62), Tn552 (36), Tn5053 (33), and Tn5090 (34). Therefore, the strategy of using an ATP binding protein to regulate target site selection may extend to the entire family. Tn5053 is particularly interesting, since it encodes three proteins which are required for its transposition: a presumptive transposase containing a D, D(35)E motif characteristic of transposases and integrases, a potential regulatory protein containing Walker A and B motifs, and a third protein of unknown function (33). Tn5053 shows some degree of target site specificity, inserting predominantly into the par locus of the conjugable plasmid RP4. It is tantalizing to speculate that the third protein of Tn5053 is a target selector, like TnsD or TnsE, directing insertions into the par locus.

The inventor's work has illustrated the role of target DNA in controlling Tn7 transposition *in vivo*, and has strongly implicated TnsC as a central player in this regulation. Single amino acid changes in TnsC can disrupt the communication between the transposon and the target site, reducing the stringency of Tn7's target site selectivity. TnsD promotes Tn7 insertion at high frequency into *attTn7*, a safe haven in the bacterial chromosome, whereas TnsE allows Tn7 access to conjugable plasmids, and thus a means to spread through bacterial populations. Avoidance of immune targets also promotes the spread of the element, rather than local hopping, and prevents one Tn7 element from inserting into another. TnsC may integrate all of these target signals, and communicate that information to the transposase.

TABLE 1

TnsC^{A225V} promotes intermolecular transposition

Tns functions	Transposition frequency
TnsABC^{wt}	$< 10^{-7}$
TnsABC^{A225V}	$8.8 (\pm 8.1) \times 10^{-6}$
TnsABC^{wt}DE	$5.5 (\pm 1.1) \times 10^{-4}$

Frequencies of transposition of miniTn7-Km^R from a high copy plasmid to pOX-G were determined using the mating-out assay, and are expressed as the number of Km^R exconjugants/total exconjugants. Each value is the average of three independent measurements.

Example 2

Materials and Methods

Media, Chemicals, and Enzymes

Luria broth (LB) and agar were prepared as described by (42). Carbenicillin and kanamycin selections were carried out at a concentration of 100μg/ml. DNA modifying and restriction enzymes were purchased from commercial sources and used according to

manufacturer's instructions. Taq polymerase was purchased from Boehringer Mannheim Biochemicals.

Bacterial Strains and Plasmids

Tn7 donor plasmids contain a miniTn7 element in which the minimal end sequences of Tn7 (Tn7L 1-166 and Tn7R 1-199) flank a selectable marker. A pBR plasmid containing a mTn7- kanamycin element with *NotI* and *SpeI* sites at the ends of the kanamycin cassette has been shown to be an effective donor. When transposition products are to be recovered by transformation, it is useful to prevent transformation of unreacted donor. One strategy is to cut the donor backbone with a restriction enzyme that does not cut within the Tn7 element or within the target DNA. Another strategy is to use donor plasmids that will not replicate with the products recovered. One strategy is to make the replication of the donor depend on a protein that is not present in the transformation strain. For example, the mTn7 element can be placed on a plasmid which does not itself encode an initiator protein for replication. A particular example is to make the donor backbone an R6K plasmid that does not encode the replicator protein pir. The R6Kpir - miniTn7 plasmid can then be grown in a strain which contains pir (supplied for example by a heterologous plasmid) and the transposition mixture transformed into a strain lacking pir. With selection for the marker on the mTn7, only insertions into the target DNA will be recovered. Subcloning Efficiency DH5alpha competent cells were purchased from GIBCO BRL and used according to the manufacturer's instructions.

The target plasmid pRM2 (SEQ ID NO:6) contains bases -342 to +165 of *attTn7* cloned into pUC18 [47]. The donor plasmid pEMΔ (SEQ ID NO:4) carries a miniTn7 element comprised of the 166 terminal bases of the left end of Tn7 and 199 bases of the right end flanking a gene conferring resistance to kanamycin [23].

Tns Proteins

The purification of TnsA and TnsB-His are described in (63). TnsA was stored in 25mM Hepes (pH 8.0), 150 mM NaCl, 1mM EDTA, 1mM DTT, 5% glycerol at -80°C. TnsB was TnsB-His, a derivative containing a C-terminal polyhistidine tag, and was stored in 25mM Hepes (pH 8.0), 500 mM KCl, 2mM DTT, 1 mg/ml BSA, 25% glycerol at -80°C. The purification of TnsC and $TnsC^{A225V}$ is a modified procedure from (24) which is described in (25) (26). Both proteins were stored in 25mM Hepes (pH 8.0), 1M NaCl, 2.5mM DTT, 1mM ATP, 10mM $MgCl_2$, 0.1mM EDTA, 10mM CHAPS, 10% glycerol at -80°C. TnsD was TnsD-His (P. Sharpe and N. Craig, in preparation), a derivative containing a C-terminal polyhistidine tag, and was purified by Ni^{+2} chromatography before being stored in 50mM Tris (pH 7.5), 2mM DTT, 500 mM KCl, 1mM EDTA and 25% glycerol at -80°C.

Transposition Reactions *in vitro*

Transposition reactions are adapted from the standard *in vitro* reaction described in (23). Reaction mixtures, 100 µl in volume, contained (final concentration) 0.25 nM pEMΔ donor, 1.9 nM pRM2 target, 26 mM Hepes, 4.2 mM Tris (pH 7.6), 50 µg/ml BSA, 100 µg/ml yeast tRNA, 2 mM ATP (pH 7.0), 2.1 mM DTT, 0.05 mM EDTA, 0.2 mM $MgCl_2$, 0.2 mM CHAPS, 28 mM NaCl, 21 mM KCl, 1.35% glycerol, 60 ng TnsA, 25 ng TnsB, either 100 ng $TnsC^{wt}$, or 100 ng $TnsC^{A225V}$, and 40 ng TnsD, unless otherwise indicated, in a 30 minute preincubation at 30°C. (TnsA=19nM, TnsB=3.1nM, TnsC=16nM, TnsD=6.5nM). Magnesium acetate was added to a final concentration of 15 mM and the reactions were allowed to proceed for an additional 60 minutes at 30°C. Products were extracted with a 1:1 mixture of phenol/chloroform, ethanol-precipitated, and resuspended in water in preparation for subsequent analyses.

PCR Primers and Amplification

Oligonucleotides used for the various PCR amplifications to analyze the products of transposition are:

NLC95 (SEQ ID NO.7): (5') - ATAATCCTTAAAAACTCCATTTCCACCCCT - (3')

NLC209 (SEQ ID NO.8): (5') - GTGATTGCACCGATCTTCTACACCGTTCC - (3')

NLC429 (SEQ ID NO.9): (5') - TTTCACCGTCATCACCGAAACGCGCGAGAC - (3')

NLC430 (SEQ ID NO.10): (5') - AATGACTTGGTTGAGTACTCACCAGTCACA - (3')

NLC431 (SEQ ID NO.11): (5') - ATGAACGAAATAGACAGATCGCTGAGATAG - (3')

NLC432 (SEQ ID NO.12): (5') - CAAGACGATAGTTACCGGATAAGGCGCAGC - (3')

Two percent of the 100 µl transposition reaction was used as the template in a given PCR amplification. 100 pg of plasmid pMCB20 was used when amplifying a marker product for size comparison on the high resolution denaturing gels. 30 temperature cycles of 94°C for 1.0 minute, 55°C for 1.5 minutes, and 72°C for 1.5 minutes were run for all amplifications, followed by a single 5 minute incubation at 72°C. The buffer composition and quantity of Taq polymerase recommended by the manufacturer (Boehringer Mannheim Biochemicals) were used for all reactions. PCR products were ethanol-precipitated, resuspended in water, and loaded on a high resolution denaturing gel.

Probe Labeling

Oligonucleotide probes were 5' end-labeled with [gamma-^{32}P] ATP substrate and bacteriophage T4 polynucleotide kinase for 45 minutes at 37°C. Labeled probes were separated from unincorporated label by size exclusion through a G50 Nick Spin Column (Pharmacia).

High Resolution Denaturing Gels

The resuspended PCR products were electrophoresed on either a 5% or 6% polyacrylamide denaturing gel and electrotransferred to Gene Screen Plus membrane (du Pont). The resulting blots were visualized by hybridization with an appropriate oligonucleotide probe at 50°C and exposed overnight to phosphorimager screens (Molecular Dynamics), which were scanned the following day.

Results

TnsC^{A225V} Supports Efficient Transposition *in vitro*

A diagram of Tn7 transposition is shown in Figure 5. Tn7 mobilizes via a cut-and-paste mechanism, whereby both ends of the element are first excised from the donor backbone by double-strand breaks, and join to the target DNA most likely via transesterification reactions to form simple insertions with short gaps at either end. Other possible intermediates of a transposition reaction are double-strand breaks (DSBs), where one end of the transposon has been excised but the other end remains attached to the donor backbone, excised linear transposons (ELTs), where both ends have been excised from the donor and neither end has joined to the target, and double-strand break, single-end joins (DSB-SEJs), where one transposon end has been broken in the donor and joined to the target molecule.

The Tn7 transposition reaction has been reconstituted *in vitro*, in which purified Tns proteins promote the transposition of a mini Tn7 element from a donor plasmid into an *attTn7*-containing target plasmid (Bainton 1993). TnsABC^{wt}+D supports this reaction with great efficiency. In the absence of TnsD, TnsABC^{wt} does not generate a detectable level of insertion products (Figure 6, lane 2) although double-strand break intermediates are seen upon prolonged incubation. By contrast, reactions containing TnsABC^{A225V} show a dramatic accumulation of simple insertions, at efficiencies that approach TnsABC^{wt}+D reactions (Figure 6, lane 5). Neither the TnsABC^{A225V} nor the TnsABC^{wt}+D reactions generate visible levels of DSB-SEJ products, indicating that the vast majority of Tn7 transposition events result in the complete (i.e., two-ended) insertion of the transposon into the target DNA, rather than a single-ended insertion event.

TnsABC+D transposition is not only efficient, it is also very target site-specific. TnsABC+D insertions occur almost exclusively into the *attTn7* site present on the target plasmid (Bainton, *et al.*, 1993, data not shown). By contrast, the TnsABC^{A225V} insertions are not limited to the *attTn7* site. Alternative restriction analysis of the TnsABC^{A225V} reaction yields a smear of products on an agarose gel (data not shown), suggestive of a population of insertions located at many different positions in the target plasmid. To

investigate the distribution of these insertions, we subjected the TnsABC^{A225V} reaction products to high-resolution analysis, as described below.

Distribution of TnsABC^{A225V}-Mediated Insertions is Highly Nonspecific

A PCR-based approach has been used to analyze insertional mutations in SV40 and yeast TRP1ARS1 minichromosomes [30, 31], and perform functional analyses of insertional mutations in yeast chromosome V and the *E. coli* supF gene [Smith, 1996 #427] and [32], respectively.

PCR was utilized to survey the distribution of TnsABC^{A225V} insertions previously seen on the agarose gel at higher resolution. The diagram in Figure 7 illustrates the PCR strategy used to amplify the population of insertion products present in a TnsABC^{A225V} reaction, with two representative insertions being shown as examples. One PCR primer (NLC95) (SEQ ID NO:7) hybridizes within the cis-acting end sequence of the inserted element and the other (NLC209) (SEQ ID NO:8) hybridizes to an arbitrary position in the target molecule. Thus, the length of the PCR product reflects the positions of the insertions into the target molecule.

Amplification of a pool of insertions generated a smear of reaction products when displayed on an agarose gel, as expected (data not shown). The PCR products were run on a 6% polyacrylamide denaturing gel to achieve single nucleotide resolution and visualized by Southern blotting and hybridization with a Tn7-specific probe (Fig. 7). The striking result is that the distribution of products is remarkably nonspecific. Insertions have occurred at nearly every base within the highly resolved lower portion of the gel. PCR products of more than roughly 200 bp in length are resolved poorly. Some areas of dense signal are seen in this region, potentially indicating preferential points of insertion. However, compression of bands could also account for the apparently singular products; analysis of these insertion products with other primer pairs supports this latter possibility (see below).

This confirms the inventor's hypothesis that the TnsABC^{A225V} machinery is capable of directing Tn7 transposition into the target plasmid with high efficiency and low specificity.

Surveillance of the Entire Target Plasmid

In the experiments above, the focus was on a relatively short region of the target plasmid pRM2 (SEQ ID NO:6). It was demonstrated that TnsABC^{A225V} can direct insertions into virtually every base pair of this region. To be certain that the phenomenon is not specific to the region of the plasmid, a family of primers was synthesized, each of which paired with a Tn7 end-specific primer to allow amplification of all regions of pRM2 (SEQ ID NO:6). These primers are spaced at approximately 500 bp intervals around the target plasmid and will amplify insertions in predominantly one orientation. Figure 8 diagrams the amplicons for each primer pair and shows a denaturing gel Southern blot of the resulting PCR products. The results indicate that the C^{A225V}-mediated insertions do occur into positions all around the target plasmid. As was seen for the original amplicon analyzed, there is considerable variability in the strength of the signal for individual points of insertion, but insertions do occur at some level at every position. Thus, the TnsABC^{A225V} machinery does not appear to have a specificity for any particular region of this target plasmid.

In another approach to investigating the possible sequence specificity of TnsABC^{A225V} target site selection, 67 independent insertions into a 12 kb plasmid were collected and analyzed. TnsABC^{A225V} transposition reactions using a target plasmid containing several *E. coli* genes were transformed into *E. coli* to select kanamycin-resistant colonies. The target plasmids were then recovered and sequenced to determine the position of each insertion. 62 out of the 63 insertions were located in different positions on the target plasmid. A comparison of the sequences of these insertions supported our previous observations that there is very little sequence specificity governing the selection of TnsABC^{A225V} target sites. Attempts to derive a consensus sequence for the 5 bp target site

duplication sequence revealed a faint preference for NYNRN (SEQ ID NO:14), but the bias is not very compelling.

Exploiting the TnsABC^{A225V} Machinery for in vitro Mutagenesis

The high efficiency and low target specificity of the TnsABC^{A225V} transposition reaction makes this a useful system for mutagenizing a variety of DNA targets. Insertional mutagenesis could be performed on cosmid libraries, cDNA libraries, PCR products, BACs, YACs, and genomic DNAs, among others. The inventor has mutagenized pUC-based plasmids, cosmids, BACs ranging in size from 5 to 120 Kb (data not shown), and *H. influenzae* genomic DNA (Gwinn *et al.*, 1997). In fact, the inventor has not encountered DNA that cannot serve as a target for TnsABC^{A225V} transposition.

Once DNA targets have been successfully mutagenized *in vitro*, the simple insertions will be recovered. For a simple insertion product to become a stable replicon, the 5' nonhomologous overhangs trailing off both ends of the inserted transposon must be removed, the gaps filled in, and the strands ligated. A simple method to perform such processing functions is to transform the pool of transposition products into a host and rely on the host's repair machinery, selecting for a transposon-borne marker. In *E. coli*, the 5' single-stranded overhangs and gaps on either end of the transposon after a simple insertion are readily repaired by the host (see below). The donor plasmid for other hosts could be customized in a number of ways to best facilitate the recovery of the desired insertional mutants.

The inventor recovered simple insertions into pRM2 in *E. coli*, since Tn7 insertions can be easily repaired in this host.

Simply transforming transposition reactions into host cells as a method to recover simple insertion isolates presents a background contributed by donor molecules that have not undergone transposition and thus continue to carry the selectable marker on a stable replicon. In order to eliminate the background false positives that can complicate a screen for insertional mutants, the ability of the unreacted donor to transform cells can be reduced. Two methods have been provided: 1) destruction of the donor plasmid's ability to replicate

by restriction digestion prior to the transposition reaction, and 2) use of a conditional replicon origin in the donor backbone which renders the donor incapable of replication in the cells being ultimately transformed with the transposition pool.

For the first method, 5 identical TnsABC^{A225V} reactions were carried out on linearized pMCB31 donor DNA paired with cosmid clone ES#3 target DNA, an approximately 50 kb replicon which contains an insert of genomic DNA from *E. tarda*. Linearizing the plasmid will prevent the donor plasmid from replicating once transformed into the host. The products were pooled for the extraction and precipitation steps, and then a portion of the resultant sample was transformed into BRL Subcloning Efficiency DH5-alpha competent cells. Assuming a 10% loss in the recovery of the DNA after the transposition reaction, the efficiency of transformation relative to µg of input donor DNA was approximately 3.8 X 10^4 colonies/µg/ml of cells. One-tenth of a microgram of donor DNA is typically used in a reaction, so by extension, if all of the product DNA from a single transposition reaction is transformed, 3800 colonies could be isolated, an efficient mutagenesis. The DH5^{alpha} cells are advertised to have a transformation efficiency of equal to or greater than 1 X 10^7 colonies/µg supercoiled pUC19/ml of cells. Simply using higher efficiency cells or electroporation cells should yield considerably higher numbers of isolates from a single transposition reaction, and probably aid in picking up rarer events.

Another method employs a heterologous origin of replication on the donor plasmid, for example, R6K. Replicons relying on this origin must be maintained in a host carrying a resident copy of the *pir* gene, which codes for the π protein, a necessary component for initiation of replication at R6K$_{gamma}$ origins. Thus, it is simple to eliminate false positives stemming from unreacted donor molecules simply by transforming the transposition products into *pir*$^-$ cells, and relying on the competent origin of replication in the target molecule for recovery of simple insertion isolates. Transposition reactions employing this donor were prepared for transformation as described above.

It is conceivable that the larger plasmid (~50kb) would be more difficult to transform after receiving an insertion because it would be a large open circular molecule approximately 10 times the size of the pRM2 (3.2 Kb) open circle with an insertion. To

gain insight into the possibility of target size limitations using the transformation method of simple insertion recovery, transpositions of the miniTn7 element from pMCB40 into the two target plasmids were directly compared. The transformation efficiencies of the two reactions were very similar. The different targets were included at comparable concentrations in the transposition reactions, but were not equimolor. The results suggest that ES#3 simple insertions can transform the cells at nearly the same efficiency as the smaller pRM2 simple insertions. It is difficult to test reaction conditions under which the cosmid target is available at the same molarity as the pRM2 target because elevated levels of total DNA in the reactions can compromise the reproducibility with which DNA is recovered after transposition.

The high transformation efficiencies demonstrate the utility of this reaction for a mutagenesis in which the simple insertion products can be stably replicated in an *E. coli* host. This same type of protocol could be used in other bacterial species and strains with development of the appropriate DNA substrates.

Discussion

TnsC^{A225V} circumvents the requirement for a targeting protein.

Tn7 demonstrates considerable diversity when it comes to target site selection. It has a sophisticated system for choosing either a highly conserved "safe haven" in the *E. coli* chromosome (*attTn7*) or somewhat random sites throughout a cell's genome or resident conjugable plasmid, mediating these different selections via alternative targeting proteins encoded by the element. In this way, Tn7 is significantly different than all other well-characterized transposable elements, whose target site selections are mediated predominantly by either the transposase alone (e.g., IS10/Tn10) or in conjunction with one other accessory protein (bacteriophage Mu). IS10/Tn10 selects a target site via a direct interaction of the Tn10 transposase with the target DNA. It has been demonstrated that particular mutations in the Tn10 transposase are capable of altering target recognition features while leaving other functions of the transposase unaffected (65). The

bacteriophage Mu, however, encodes a transposase, MuA, and an ATP-dependent activator of MuA, MuB. MuB functions as an accessory protein that, when complexed with target DNA, attracts the MuA transposase to the site of insertion. It is likely that having more proteins involved has allowed Tn7 to be more adaptive to environmental changes when choosing its new sites of residence, and ensured its survival by enabling it to employ a more tailored approach to disseminating itself amongst various cell populations.

This example has focused on the role of TnsC in the selection of a target site. As discussed, TnsC has been implicated as the major communicator between the TnsAB transposase bound to donor DNA, and the TnsD or TnsE targeting proteins, complexed with target DNA. Experiments have shown that TnsC does have the capacity to bind DNA nonspecifically in the absence of TnsD and TnsE (ref) but attempts to isolate simple insertions *in vivo* and *in vitro* in the absence of the targeting proteins proved unsuccessful with wild-type TnsC (23). Isolation of the $TnsC^{A225V}$ mutant, however, has permitted the inventor to circumvent this requirement and isolate simple insertions from reactions lacking TnsD and TnsE. Not only does the mutant facilitate the recovery of simple insertions, it does so very efficiently.

Ability of $TnsC^{A225V}$ to Insert Nonspecifically

It is clear that $TnsC^{A225V}$ has a considerable gain of function over wildtype TnsC, as evidenced by the increased yield of simple insertions in a standard *in vitro* reaction (25) (26) (this example). A more detailed evaluation was necessary to determine the actual sites of insertion because restriction digests of the product pools indicated that there is extensive variability in site selection relative to TnsD-mediated insertions, which are targeted almost exclusively to the attachment site. PCR amplification of pools of transposition products followed by high resolution denaturing gel analysis of several independent reactions has revealed that the insertions into the pRM2 target plasmid are detectable at every base visible within the well-resolved portions of the gels. Although the target site selection is not completely random (there are differences in band intensities), one possibility is that the

nonspecific DNA binding activity of TnsC has been enhanced in the $TnsC^{A225V}$ mutant, giving the protein the capacity to direct the TnsAB transposase to the wide variety of insertion sites observed.

It is possible that the $TnsC^{A225V}$ mutation has altered TnsC in such a way that it simulates a TnsC-TnsE complex, capable of insertions at more random sites. Perhaps the role of TnsE is to strengthen TnsC's nonspecific interaction with the target DNA, thereby promoting insertions into sites where TnsC and TnsE happen to complex. The ability of TnsE to preferentially direct transposition to conjugating plasmids (14) holds true when $TnsC^{A225V}$ is substituted for wild-type TnsC (SEQ ID NOS:1 and 2) (25) (26). This suggests that this mutation in TnsC does not compensate for all specific activities of a targeting protein. The $TnsABC^{A225V}$ reaction is also sensitive to the presence of the target site specific protein TnsD, as evidenced by a detectable increase in the frequency of insertions when TnsD is present.

These observations may explain why Tn7 has chosen to preserve a more complicated target site selection mechanism. In a cell containing only wildtype proteins, an extra layer of regulation can be exercised when two proteins complex to direct insertions, and the result may be less deleterious to cell populations than the somewhat rampant levels of insertions observed in reactions with the $TnsC^{A225V}$ in the absence of targeting proteins. Occurrence of a mutation like $TnsC^{A225V}$ in nature would decrease the specificity and increase the frequency of insertions, the consequence of which could quite possibly be more insertions into essential genes.

It is conceivable that TnsC has always played the primary role in directing the TnsAB transposase to insert, and the targeting proteins are more accessory. The inventor has envisioned TnsD binding DNA near the attachment site, and TnsC acting as an activation bridge to the transposase, but an alternative view is that the ability of TnsC to bind DNA plays a more central role in directing the donor complex to an insertion site, and TnsD has the role of "steering" a TnsAB+TnsC complex to a particular point of insertion. The A225V point mutation could confer the ability for TnsC to "steer" the donor complex to insert without the aid of a target-binding protein.

There is No Apparent Sequence Preference at the Point of Insertion

Two main approaches have been taken herein to analyze the $TnsC^{A225V}$-mediated insertions at nucleotide resolution. The first involves scanning along a short segment of DNA using PCR and high resolution denaturing gel analysis, quantitating specific signals at each base in a processive manner, and attempting to flush out a sequence motif common to those with the highest signals or lowest signals. The second method focuses on the recovery of the more frequent insertions only, those that can be recovered by simply transforming the transposition products, and relying on the host to conduct a successful repair of the replicons. These two methods provide different views of a common process. Since recovery of specific insertions is reliant of the process of transformation, rare insertion events that can be visualized by the PCR/denaturing gel method will most likely be severely underrepresented in a population of recovered transformants, if we assume that a higher concentration of a specific template will give rise to a diagnostic PCR product of higher intensity. This should bias the representative data accumulated from transposition product transformations to overlap with the subset of PCR products analyzed by denaturing gel analysis with the highest band intensities. In this way, both types of data are valid for attempting to determine a preferred insertion site.

The inventor's search for a common insertion site motif failed to uncover any preferred single nucleotides or groups of nucleotides that showed a higher incidence amongst the most intense signals on a denaturing gel or amongst the insertions isolated by transformation. Similarly, there were no apparent motifs common amongst the least preferred sites analyzed in the denaturing gel analysis. The lack of a sequence preference for insertions with this reaction is a very desirable result if it is to be employed as a highly nonspecific method for mutagenizing DNAs.

TnsC^{A225V}: A Tool for *in vitro* Mutagenesis

The impressive efficiency and low specificity of the TnsABC^{A225V} *in vitro* reaction makes the reaction an excellent tool for *in vitro* mutagenesis. The high efficiency of the reaction (i.e., the high percentage conversion of donor substrate to double-ended simple insertions) is critical when considering how the recombinant DNAs will be recovered. The observation that the majority of the molecules resulting from a reaction that contain a junction between the donor DNA and the target DNA are double-ended simple insertions provides an advantage over alternative transposon-based insertional mutagenesis systems because large portions of the junctions seen in these reactions can be single-end joins (Rowland, S. J. *et al. EMBOJ. 14*:196-205 (1995)). This study has demonstrated that standard commercially available *E. coli* competent cells are capable of repairing the characteristic gapped molecules formed as a result of a Tn7 simple insertion, provided the target DNA contains an origin capable of replicating in *E. coli*. Thousands of isolates can be recovered from a single transposition reaction starting with sub-microgram quantities of donor and target DNA. High efficiency cells should yield even greater numbers of isolates. Tn7 insertional mutants could be recovered from many different organisms as long as the target DNA carries information required to replicate in its respective host, the gaps can be repaired by the host, and DNAs can be reintroduced into the host with reasonable efficiency.

Cosmid clones have been successfully mutagenized and recovered by the method just described. Pilot reactions were done using purified cosmid clones. But it would be very simple to mutagenize an entire cosmid library and select for mutants by the same process. Replicons as large as 125 kb (a BAC, data not shown) have been successfully targeted and recovered. An earlier study of the inventor demonstrated that the ability of transposition machinery to recognize whether or not a potential target molecule already contains an insertion breaks down as the distance between two insertion sites increases (52). It has been shown that the degree to which a target molecule is "immune" to a second insertion has an inverse relationship to the length of separation of the sites of insertion. TnsC^{A225V} has demonstrated a sensitivity to immunity signals. To date, the inventor has

seen very few examples of double insertions into plasmids in the 40-50 kb range, suggesting that this tool will be highly effective for mutagenizing cosmids or plasmids in the 1-50 kb range.

Example 3

A Kit For Making Transposon Insertions

The kit provides transposon insertions into DNA *in vitro*. These insertions can be used to provide priming sites for DNA sequence determination, or to provide mutations suitable for genetic analysis, or both.

Section A: Reaction Constituents

A1) PROTEINS

TnsA 30 μg/ml in 10% glycerol

TnsB 20 μg/ml in 25% glycerol

$TnsC_{127}$ 100 μg/ml in 10% glycerol

Proteins were kept at -70°C.

A2) BUFFER CONSTITUENTS

HEPES	0.25 M pH 8.1
Tris[C1]	0.25 M pH 7.6 [can be omitted]
BSA	10 mg/ml
tRNA	50 μg/ml [can be omitted]
DTT	1 M

ATP	100 mM
MgAcetate	375 mM

A3) TRANSPOSON DONOR PLASMID

100 µg/ml

The essential features of the plasmid are described above as containing the R6K conditional replicon.

A4) CONTROL TARGET PLASMID

pLITMUS28 400 µg/ml

This plasmid contains both pUc and Mi3 origins, a lacZ' MCS and amp. See the Figure legend for figure 10B.

(New England BioLabs, 32 Tozer Road, Beverly, Massachusetts, 01915)

A5) SEQUENCING PRIMERS

NLC94 (SEQ ID NO:13)	3 pmol/µl
NLC95 (SEQ ID NO:7)	3 pmol/µl

Section B: (Can be supplied by user)

B 1) FOR THE REACTION *in vitro*

water; Millicue or equivalent recommended

Target DNA not carrying Kanamycin resistance (0.4 - 0.5 µg per reaction)

Water bath or heat block, 30°C

1.5 ml microtubes or other vessel; one per reaction.

B2) FOR STOPPING THE REACTION

when using chemically competent cells

Water bath or heat block, 75°C. **Note: not 65^0C.**

when using electocompetent cells

Distilled phenol equilibrated with TE or Tris pH 8.0

Chloroform equilibrated with TE or Tris pH 8.0

EtOH for precipitation

NaAcetate 3 M

Water or 1 mM Tris pH 8 or TE

B3) FOR RECOVERING INSERTIONS:

B3a) Transformable cells:

Any standard *E. coli* strain can be used; we have used ER1821, ER2502 and MC1061 (New England BioLabs, 32 Tozer Road, Beverly, Massachusetts, 01915).

Any kanamycin-sensitive organism in which *npt* can be expressed can also be used with the KanR donor, including but not limited to, *Salmonella,* other enteric organisms, *Haemophilus, Rhizobium,* and *Bacillus.* With a suitably altered selectable marker on the transposon donor plasmid, any prokaryotic or eukaryotic organism into which exogenous DNA can be introduced, may be used to recover insertions.

In this example,

B3ai) Chemically competent ER1821 New England BioLabs, 32 Tozer Road, Beverly, Massachusetts, 01915 (2×10^7 transformants/μg of LITMUS or similar plasmid) was used. A sample protocol for preparing these is provided below, section D1

In example 2 we show the use of

B3aii) Electrocompetent MC1061 (ATCC# 53338)

(7×10^9 transformants/μg of pLITMUS-28 or similar plasmid). A sample protocol for preparing these is provided below, section D2

Commercially available competent or electrocompetent cells may also be used. The method of determining competence of these preparations is provided below, section D3.

B3b) Outgrowth media:

Rich Broth (D4a below) or mSOC (D4d) without drug, or equivalent.

0.4 ml per reaction; we recommend three reactions as a standard pilot experiment (see Section C below).

B3c) Selective media:

Rich Agar with drug (D4b), or equivalent

at least 1 plate per reaction; the standard pilot experiment described in section C require 6 plates, three with two drugs and three with one drug.

Kanamycin is REQUIRED to select for the transposon of the present example

Ampicillin is used for the RECOMMENDED positive control. Carbenicillin can be substituted

For the example of Section C, below, RB Kan Amp (3 plates) and RB Amp only (3 plates) are used. If the target plasmid carries some other drug resistance, the experimental reaction in the pilot experiment should be plated on Kanamycin plus that drug.

B4) FOR DNA PREPARATION FOR SEQUENCING (see example 2):

Any standard procedure that ordinarily gives sequencing grade DNA. We have tested Qiagen spin columns and gravity flow plasmid preparations.

Section C. Tn7 *in vitro* transposition reaction protocol

C1. REACTION VOLUME = 100µl

C2. RECOMMENDED PILOT EXPERIMENT 3 samples to be carried through.

Tube 1	Experimental (Target DNA, protein and donor plasmid added)
Tube 2	Reaction positive control (pLITMUS28, protein and donor plasmid added)
Tube 3	Reaction negative control (Target DNA added, no protein, donor added)
Tube 2	is also used as a transformation positive control

In this example, all tubes have pLITMUS28 as target (tubes 1 and 2 are duplicates). Tube 2 need not necessarily be included in every experiment.

C3. MAKE UP a mix using reagents of Section A:

per reaction:

(73.9 µl H_2O) - (volume of target DNA); in this example, target DNA is 1 µl

10 µl	Hepes (250mM pH 8.1)
1 µl	Tris (250mM pH 7.6)
0.5 µl	BSA (10 mg/ml)
2.1 µl	tRNA (50µg/ml)
0.2 µl	DTT (1M)
2 µl	ATP(100mM)

C4. DISPENSE mix of step 3 to each tube (89.7µl) - (volume of target DNA)/reaction; in this example, this is 88.7µl.

C5. ADD target DNA of section B (0.4μg) to tubes 1-3. In this example, this is pLITMUS28, 1μl. This works well for plasmid targets. For cosmids, 0.5μg worked well when the cosmid was around 10 times the size of the donor (5.2kb) i.e. a molar ratio of around 2: 1 (donor to target). Increasing the ratio to 4:1 *decreased* the efficiency slightly.

C6. ADD to each tube

	Tube 1	Tube 2	Tube 3
TnsA	1.3 μl	1.3 μl (40ng)	0
TnsB	3 μl	3 μl (20ng)	0
$TnsC_{127}$	1 μl	1 μl (100ng)	0
dH2O	0	0	5.3 μl

C7. ADD 1μl donor DNA (0.1μg pMCB40). Mix well by pipetting up and down a few times.

	Tube 1	Tube 2	Tube 3
Donor pMCB40	1 μl	1 μl	1 μl

C8. INCUBATE 10 minutes at 30°C (assembly reaction)

C9. ADD 4 μl MgAc (375mM) to each tube. Mix well by pipetting up and down a few times

	Tube 1	Tube 2	Tube 3
MgAc	4 μl	4 μl	4 μl

C10. INCUBATE 1 hour 30°C (transposition reaction)

C11. HEAT INACTIVATE 75°C 10 minutes. **Note: 65°C is <u>not</u>** adequate.

C12. TRANSFORM using chemically competent cells (see procedure of section D1):

a. Add 10 µl of the reaction mix to 100 µl competent cells thawed on ice.

b. Incubate 1 h on ice.

c. Heat at 37°C for 45 sec.

d. Chill on ice 2 min.

e. Dilute the transformation mix into 0.4 ml RB (total volume 0.5 ml).

f. incubate 40 min at 37^0C.

g. plate 100 µl tubes 1-3 on Kanamycin-containing selective media.

h. plate dilutions of tube 2 on medium selective for the target plasmid only: dilute 100 fold (10 µl/1 ml) and 1000-fold (1 µl/1ml) and plate 100 µl of undiluted and of each dilution (3 plates)

In this example, selective medium was RB Kan (20 µg/ml) Amp (100 µg/ml) (tubes 1-3) and RB Amp (100 µg/ml, tube 2). Competent cells were ER1821, chemically competent (Section Dl).

C13. Transformation result:

On Kan Amp:

Tube 1 285 colonies

Tube 2 600 colonies

Tube 3 0 colonies

On Amp only:

Tube 2 confluent (undiluted)

Section D: Recipes and auxiliary procedures

D1) Chemically competent cells (*E. coli*):

a. Inoculate a single colony from an RB agar plate (see D4b) into 2 ml of RB (D4a) in a plating tube. Shake overnight at 37°C.

b. Subculture the overnight 1:100 in 1 Volume Unit of RB+20 mM $MgSO_4$ (typically 250 ml). Grow to OD_{590}=0.4-0.6 or Klett=60 (~2-3 h).

c. Centrifuge 5,000 rpm 5 min at 4°C.

d. Gently resuspend pellet in 1/2.5 Volume Unit **ice cold** TFBI (see below, D4f). **Keep all steps on ice and chill all pipets, tubes, flasks, etc. from this point on.**

e. Incubate on ice for 5 min.

f. Centrifuge 5,000 rpm 5 min 4°C

g. Gently resuspend pellet in 1/25 original volume cold TFB2 (D4g). For 250 ml of original subculture, use 10 ml TFB2.

h. Incubate on ice 15-60 min. before aliquoting 100 µl/tube for storage at -70^{0}C. Quick-freeze the tubes.

i. To transform, thaw an aliquot on ice; add DNA; incubate 1 h on ice; heat shock 45 seconds at 37°C; incubate on ice 2 min; dilute 5-fold into RB with no drug (for

phenotypic expression); grow with vigorous aeration at 37°C for 20 min.; plate on selective medium.

This procedure works with most strains and should routinely give $>10^7$ cfu/μg of pLITMUS28 (using 0.1 ng/transformation). Frozen cells last at least a year.

D2) Electrocompetent cells (*E. coli*)

D2a. Rationale and comments

This procedure prepares cells for use in gene transfer employing an electroporator device such as that supplied by BioRad. DNA is introduced into cells by means of an electric field.

Successful electroporation requires a low electrolyte concentration, to avoid arcing (and cell killing) in the device. Cells are grown to midexponential phase, washed extensively in distilled water and sterile 10% glycerol, concentrated 500-fold in glycerol, aliquoted and stored at -70°C.

Any strain can be used for this purpose, although some strains are said to give larger numbers of transformants. Resuspended cells should be well-dispersed for best results. Some strains resuspend more evenly in the low electrolyte solutions; some lyse under these conditions with rough treatment.

The electroporation procedure itself involves transfer of the thawed cells to an electroporation cuvette (which has leads that contact the device appropriately), addition of DNA, imposition of the electric field, recovery from this treatment (by incubation in broth), and plating selectively.

Efficiency of transformation with this method is 100-500 fold greater than with standard transformation. It is therefore especially suitable when low transformation efficiency is expected or large numbers of transformants are desired. The method is said to be especially suitable for introduction of large DNA molecules.

D2b. Preparation of electrocompetent *E. coli* cells (from BioRad recommended procedure)

i. Materials for 2 ml of electrocompetent cells (20 aliquots, 100 µl):

overnight culture of desired strain (in Rich Broth (D4a) or Luria Broth (D4c))	1 ml
Luria Broth (D4c)	1 L
dH_2O, sterile, 4°C or 0°C	1.5 L
10% (w/v) glycerol, sterile (D4h)	22 ml
1 L sidearm flasks	2
250 ml centrifuge bottles	6
50 ml Oak Ridge centrifuge tubes	2
1.5 ml microtubes, polypropylene	20
Pipet tips (sterile) for P200 or equivalent	20
Sterile glass or plastic pipets, 25 ml	3
Klett-Summerson colorimeter	
High speed centrifuge (e.g. Beckman J21)	
Micropipetter, e.g. Gilson Pipetman P200	
Water bath rack that can be used to immerse tubes in liquid nitrogen.	
Liquid nitrogen bath for quick freezing	

ii. Procedure for making electrocompetent cells

Be sure the sterile dH_2O and 10% glycerol is cold.

If necessary, distribute the Luria Broth to sidearm flasks, 500 ml/flask

Inoculate each flask with 0.5 ml of the overnight culture

Incubate with shaking until Klett=90 (5 x 10^8 cfu/ml). Quick conversion if Klett is not available: 1 OD=150 Klett Units; 10^9 cells/1.1 OD)

Chill on ice with swirling, until cold. **It is very important to keep everything cold from this point on.**

Transfer to centrifuge bottles, 167 ml/bottle or as desired.

Centrifuge 4,000 rpm 15 min 5°C in JA14 rotor in Beckman. Decant supernatant.

Resuspend gently in equal volume (1 L total) cold sterile distilled water. Keep in an ice bath while resuspending. Repeated pipetting will help; chill pipets for this use.

MC1061 cells (ER1709) can be kept on ice at this stage for at least an hour

Centrifuge 4,000 rpm 15 min 5°C in JA14 rotor in Beckman; decant supernatant.

Resuspend gently in 1/2 volume cold sterile distilled water (0.5 L total). Keep in an ice bath while resuspending. Cells can now be combined into three bottles if desired.

Centrifuge 4,000 rpm 15 min 5°C in JA14 rotor in Beckman. Decant supernatant.

Resuspend in 1/50th volume cold sterile 10% glycerol (20 ml total). Keep cold while resuspending.

Transfer entire amount to a 50 ml Oak Ridge tube (35 ml capacity).

Centrifuge 4,000 rpm 15 min 5°C in JA17 rotor in Beckman, with balance tube. Decant supernatant

Resuspend in 1/500th volume (2 ml total) cold 10% glycerol. Keep cold.

Distribute 100 µl/tube to microtubes in ice water bath rack; immerse rack in liquid N_2; transfer to box; store at -70°C

D2c. Procedure for electroporation of poratable *E. coli* cells (from BioRad recommended procedure)

D2ci. Materials (per electroporation reaction)

Electrocompetent cells	100 µl
18 x 150 mm culture tubes	1
Electroporation cuvettes (BioRad cat#1652086 or equivalent)	1
mSOC (see D4d)	1 ml
Pasteur pipets, sterile	1

DNA to be transformed; in low ionic strength medium, e.g. dH_2O or TE (see D4i).

Electroporator (BioRad Gene Pulser or equivalent)

Ice bath trays for cuvettes and outgrowth tubes

Rollordrum in 37°C incubator or other means of incubating culture tubes

Selective agar plates and plating materials

37°C or suitable temperature incubator

D2cii. Procedure

Be sure all materials are set up ready to go before getting cells out of the freezer. The DNA must be added and the electorporation done as soon as the cells are thawed; cells will lyse after a short time, resulting in arcing as the medium becomes more conductive.

Chill cuvettes and hold on ice (>5 min). Transformation efficiency declines at least 100-fold if cuvettes are at room temperature

Set BioRad Gene Pulser to 25 μF capacitance, 2.5 kV, and the pulse controller to 200 Ω (maximum voltage)

Thaw electrocompetent cells at room temperature and transfer to ice.

In a cuvette mix 40 μl cells with 0.4 pg-0.3 μg DNA. Shake the suspension to the bottom of the cuvette, rap on table to shake loose air bubbles.

Place the cuvette in the holder

Apply one pulse by pushing both red buttons until a beep is heard. This will result in a pulse of 125 kV/cm with a time constant of 4-5 sec.

Immediately add 1 ml mSOC to the cuvette and gently but quickly resuspend the cells. A P1000 with sterile blue tips or sterile pasteur pipets can be used for this. A 1 min delay in adding the medium results in 3 fold decrease in transformation efficiency.

Transfer cells to culture tube.

Incubate 37°C 1 hour

Plate on selective media.

D3) Standardization of transformation or electroporation

D3a. Rationale and comments

To ensure that gene transfer is successful, we recommend that the cells prepared above (D1 or D2) or purchased commercially be transformed with a standard DNA dilution series before experimental use. Below is an example of such a standardization for electrocompetent cells (D2). Chemically competent cells will yield 100-500 fold fewer transformants, so dilutions given below should be appropriately adjusted.

D3b. Materials for a standardization experiment

Dilutions of standard DNA, usually a high-copy small plasmid (e.g. LITMUS28), in TE:

A	1	ng/µl
B	10	pg/µl
C	1	pg/µl
D	100	fg/µl

Selective agar plates; RB 1.5% Amp 100 µg/ml for pLITMUS28	12
Dilution medium, usually 0.85% saline	7ml
Dilution tubes, usually 13 x 100 mm	7
Sterile plastic or glass pipets, 0.1 ml	10
Sterile plastic or glass pipets, 0.2 ml	1
Sterile plastic or glass pipets, 1 ml	1

Micropipetters, e.g. P200 and P20 or P10, for DNA transfer and dilution

series

Pipet tips for P200 and P20 or P10

Spreader

Ethanol or isopropanol for flaming the spreader

37°C incubator

D3c. Procedure for standardization experiment

D3ci. Set up dilution tubes below and label plates beforehand or while cultures are growing out.

D3cii. Carry out electroporation as above (D2) with DNA dilutions A-D

D3ciii. Place cultures on ice to prevent further growth while making dilutions and plating as below.

D3civ. Dilute in saline:

Sample A	10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}
Sample B	10^{-1}, 10^{-2}
Sample C	10^{-1}
Sample D	no dilutions

This can be carried out as:

10^{-1} dilution: 100 µl sample + 900 µl saline

10^{-2} dilution: 10 µl sample + 1 ml saline

10^{-3} dilution: 10 µl 10^{-1} dilution + 1 ml saline

10^{-4} dilution: 10 µl 10^{-2} dilution + 1 ml saline.

D3cv. Plate on selective media by spreading; flame the spreader after each plate:

Dilutions:	undiluted	10^{-1}	10^{-2}	10^{-3}	10^{-4}
Samples:					
A		0.1 ml	0.1 ml	0.1 ml	0.1 ml
B	0.1 ml	0.1 ml	0.1 ml		
C	0.1 ml	0.1 ml			
D	0.1 ml				
	0.2 ml				
	0.5 ml				

D3vi. Example of result:

Sample	DNA added	Dilution/vol plated	Colonies	Transformants per ml	per µg
A	1 ng	1/0.1	Confluent		
		2/0.1	very numerous		
		3/0.1	~1000		
		4/0.1	71	7×10^6	7×10^9
B	10pg	0/0.1	very numerous		
		1/0.1	405		
		2/0.1	49	4×10^4	4×10^9
C	1 pg	0/0.1	very numerous		
		1/0.1	106	1×10^4	1.1×10^{10}
D	100 fg	0/0.1	~500		
		0/0.2	173		

0/0.5	75	8×10^2	8×10^9

Average transformants/μg 7.6×10^9

D4) Recipes

Bacteriological

D4a) RB, per liter

Tryptone (Difco)	10 g
Yeast Extract (Difco)	5 g
NaCl	5 g
NaOH (1 N)	2 ml
Autoclave	

D4b) RB Agar with drug, per liter

Tryptone (Difco)	10 g
Yeast Extract (Difco)	5 g
NaCl (Baker)	5 g
NaOH (1 N)	2 ml
Agar (Difco)	15 g
Autoclave	

Drugs: add after autoclaving and cooling to 55°C, per liter:

Kanamycin (REQUIRED)	20 mg

Other drugs that MAY be added, per liter; choice depends on target plasmid:

Ampicillin or carbenicillin	100 mg

Chloramphenicol	15 mg
Tetracycline	15 mg

Others drugs not tested but presumably usable in an appropriate host strain:

- Spectinomycin
- Streptomycin
- Gentamycin
- Erythromycin
- Rifampicin (recessive marker)
- Bleomycin
- Other antibacterial small molecules

D4c) Luria Broth, per liter

Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
$MgCl2{\cdot}6H_2O$	1 g
glucose	1 g

Aliquot and autoclave. For preparing electrocompetent cells (C2) it is convenient to aliquot 500 ml/flask in 1 L sidearm flasks before autoclaving.

D4d) mSOC, per liter (modified from BioRad recipe)

Luria Broth	1	L
$MgSO_4$, 1M sterile	10	ml
40% glucose, sterile	6.5	ml

Add $MgSO_4$ and glucose sterilely to sterile Luria Broth

D4e) 0.85% saline, per liter

NaCl	8.5	g

Distribute in suitable aliquots, autoclave.

Buffers and storage media

D4f). TFBI

30 mM KOAc (potassium acetate)

100 mM RbCl

10 mM $CaCl_2$

50 mM $MnCl_2$

15% glycerol

Adjust to pH 5.8 with acetic acid and filter to sterilize. It is convenient to make this as:

5 g RbCl (Alfa)

12.3 ml KOAc	1 M	
4.1 ml $CaCl_2$	1 M	
20.5 ml $MnCl_2$	1 M	(this is pink)

61.5 g glycerol; pH to 5.8 with ≤ 8 ml HOAc 0.1 M

make up to 410 ml; distribute in 100 ml sterile aliquots; and use 1 aliquot/250 ml culture.

D4g). TFBII

10 mM MOPS

75 mM $CaCl_2$

10 mM RbCl

15% glycerol

Adjust pH to 6.5 with KOH and filter to sterilize

Make up as

1.5 ml MOPS 1 M pH 6.5 (this is yellow)

11.25 ml $CaCl_2$ 1 M

1.5 ml RbCl 1 M

22.5 g glycerol

pH with 1 N KOH; make to 150 ml, filter; use 10 ml per original 250 ml culture.

D4h) 10% glycerol, per liter

Glycerol	100	g
dH_2O	1	L

Aliquot and autoclave

D4i) TE, per liter

1 M Tris pH 8.0	10 ml
0.5 M EDTA pH 8.0	2 ml

Example 4

Random Insertion of Primers For Sequencing

Section A: Components used for transposition reaction

A1) PROTEINS

TnsA 40 µg/ml in 10% glycerol

TnsB 20 μg/ml in 50% glycerol

$TnsC_{127}$ 100 μg/ml in 50% glycerol

Stored at -70^{0}C.

A2) BUFFER CONSTITUENTS

HEPES	0.25 M pH 8.1
Tris[Cl]	0.25 M pH 7.6 [can be omitted]
BSA	10 mg/ml
tRNA	50 μg/ml
DTT	1 M
ATP pH7	100 mM
MgAcetate	375 mM
TnsD storage buffer	TnsD is stored in the following buffer: 3.3μl 500 mM KCl, 50 mM Tris-HCl (pH 8.0), 1 mM EDTA, 2 mM DTT and 25% glycerol

A3) TRANSPOSON DONOR PLASMID

pEM delta R.adj to 1 50 μg/ml

(Sequence appears in Figure 9B and SEQ ID NO:3))

A4) TARGET PLASMID

1) pER183 mini-cleared lysate 200 μg/ml

2) pER183 CsCl preparation 400 μg/ml

3) pRM2 400 μg/ml

(Sequence of pER183 appears in Figure 10A and SEQ ID NO:5)

Section B: Components used for processing reaction

Phenol/chloroform equilibrated with TE

Phenol equilibrated with Tris pH 8.0

NaAcetate 3 M

Ethanol (EtOH)

BstEII New England BioLabs, 32 Tozer Road, Beverly, Massachusetts, 01915

DNA Polymerase I Holoenzyme New England BioLabs, 32 Tozer Road, Beverly, Massachusetts, 01915

T4 DNA Ligase New England BioLabs, 32 Tozer Road, Beverly, Massachusetts, 01915

10X Fi/L buffer (section I3)

10X Buffer 3 (NEB#007-3) New England BioLabs, 32 Tozer Road, Beverly, Massachusetts, 01915

tRNA 1 mg/ml

DNA buffer (section I2)

TE (section I4)

Section C: Components used for recovery of insertions

MC1061 electrocompetent cells (made and used as in Example 3, D2 and D3)

Selective media (made and used as in Example 3, D3 and D4)

Section D: Components used for sequence determination

Dl) SEQUENCING PRIMERS

NLC94 3.2 pmol/µl.

Sequence of this primer (SEQ ID NO:13):

5' AAAGTCCAGTATGCTTTTTCACAGCATAAC

NLC95 3.2 pmol/µl

Sequence of this primer (SEQ ID NO:7)

5' ATAATCCTTAAAAACTCCATTTCCACCCCT

D2) QIAPREP SPIN MINIPREP KIT (Qiagen Cat # 27106)

D3) ABI Sequencer (info) and reagents

Section E: *in vitro* transposition protocol

E1) MAKE UP Mix:

208.2	µl	dH_2O
30	µl	Hepes (250 mM pH 8.1)
3	µl	Tris (250 mM pH 7.6)
1.5	µl	BSA (10 mg/ml)
6.3	µl	tRNA (50µg/ml)
0.6	µl	DTT (1M)
6	µl	ATP (100mM)

E2) DISPENSE 85.2 µl to three tubes

E3) ADD target DNA of A4, 2 µl

E4) ADD to each tube

	Tube 1	Tube 2	Tube 3
TnsA	1.3 µl	1.3 µl	1.3 µl
TnsB	1 µl	1 µl	1 µl
$TnsC_{127}$	1 µl	1 µl	1 µl
D buffer	3.3	3.3	3.3 µl
Donor	2	2	2 µl

E5) INCUBATE 30 minutes at 30°C (assembly reaction)

E6) ADD 4 µl MgAc (375mM) to each tube.

E7) INCUBATE 1 hour at 30°C (insertion)

Section F: Reaction processing

In this example, the transposon donor was capable of replicating in the host used for recovery of insertions. Transformation of the reaction mixture on plates selecting for the transposon and the target markers might well result in many colonies with two different plasmids, rather than with a single plasmid containing both markers. For this reason, we digested the reaction with a restriction endonuclease cleaving in the donor replicon but not within the transposon or in the target DNA. In addition, we examined the consequences of repairing the strands not ligated by the transposition reaction, using DNA polymerase I holoenzyme and ligase.

Per reaction (100 µl):

PC extract:

Add 100 µl phenol/chloroform, vortex

Centrifuge 5' in microfuge

Backextract

Remove organic phase to a new tube with 100 µl TE; vortex

Centrifuge 5' in microfuge

Combine aqueous phases (185 µl total)

EtOH precipitate

20 µl 3 M NaAc

500 µl EtOH

chill on dry ice

Centrifuge 5 min in microfuge

Drain supernatant, air dry

Resuspend in 100 µl DNA buffer

Divide each reaction for further treatment (all volumes are µl)

Treatment:	Repair Digestion A	Diget B
1) Repair/ligation		
DNA	40	40
10X Fi/L	5	-
dH_2O	2	-
Pol I (10,000 µ/ml)	2	-
a) Incubate 15 min room temperature		
Ligase (400,000 u/ml)	1	-
b) Incubate 4 h 16°C		

2) Digestion

1 M NaCl	6.0	-
10 X buffer 3		6.0
BstEII (10,000 u/ml)	1	1

Incubate 60°C 1 h

3) Protein removal, buffer exchange 1

Phenol, equilibrated	50	50

a) Mix, centrifuge 5' in microfuge

b) Back extract organic phase with DNA buffer

c) Combine aqueous phase

Total volume, step 3c	100	100
3 M NaAc	10	10
tRNA 1 mg/ml	1	1
EtOH	120	120

a) Incubate 5 min room temperature

b) Centrifuge, discard supernatant

c) Wash twice with cold 70% EtOH (100 µl)

DNA buffer	50	50

d) Resuspend

Final volume, step 3d	50	50

4) Buffer exchange 2		
Re-precipitation		
DNA from step 3d	35	35
3 M NaAc	5	5
EtOH	137.5	137.5

a) Incubate -70°C overnight

b) Centrifuge, discard supernatant

c) wash twice 200 µl 70% EtOH

TE	50	50

d) Resuspend

Section G: Recovery of insertions

Electroporated 10 µl of samples into MC1061 following procedure of Example 3, section D3

Table 2 Sample codes, treatments, and target concentrations corrected for losses during manipulation

	Target Name	Treatment	Target Selection	[Target DNA] (fmol/µl)
1A	pER183	Fi/L, Dig	Cam	0.015
1B	pER183	Digested	Cam	0.05
2A	pER183	Fi/L, Dig	Cam	0.98
2B	pER183	Digested	Cam	0.42
3A	pRM2	Fi/L, Dig	Amp	0.66
3B	pRM2	Digested	Amp	0.56

Table 3: Colony forming units per ml on appropriate selective plates

Sample	1A	1B	2A	2B	3A	3B
Donor (or recomb)						
Kan Only	130	1.8×10^5	5×10^3	7×10^3	3.7×10^4	4×10^4
Recipient						
Cam only	1×10^4	8×10^5	4×10^4	6×10^6		
Amp only					3×10^7	4×10^7
Colonies/fmol	6×10^4	1.6×10^6	4×10^3	1.4×10^6	4.5×10^6	7×10^6
Recombinant						
Kan Cam	16	2.7×10^3	880	1×10^4		
Kan Amp					1.1×10^5	4×10^4
Recomb/recip	1×10^{-3}	3×10^{-3}	2×10^{-2}	1×10^{-3}	4×10^{-3}	1×10^{-3}

75 recombinant colonies were chosen, 31 from samples 2A, 44 from samples 2B, for further characterization

H. Determination of sequence location.

1. Procedure summary

75 recombinant colonies were picked into 0.5 ml RB in racked array for storage. Subcultures of these storage cultures were grown with selection (RB Cam Kan), and minipreps made according to the directions of the manufacturer for large plasmids of low copy number.

DNA concentration of the plasmid preps was determined by comparison with a dilution series of linearized pLITMUS28 on agarose gels. Plasmid preps were linearized for this purpose with an enzyme that cleaves once in the target plasmid and not in the transposon (SacII).

Primers NLC94 (SEQ ID NO:13) and NLC95 (SEQ ID NO:7) were used for sequence determination, using flourescently-labeled dideoxynucleotide sequencing reagents from Applied Biosystems.

Sequences were run on an ABI sequencer, and sequence acquisition, editing and assembly was carried out with the supplied programs (SEQED, FACTURA and AUTOASSEMBLE).

Output is Fig. 11

2. Results

a. Table 4: Summary result of 75 recombinants (CamR KanR colonies), potential Tn7 insertions into pER183.

Total DNA preps	75
DNA concentration too low to attempt sequence:	7

Transformant contained two plasmids, not sequenced:	1	
Total not sequenced	8	
DNA preps sequence attempted		67
Sequence unreadable (miscellaneous reasons)	2	
Sequence unreadable because 2 insertions in one plasmid	1	
Total sequence unreadable	3	
DNA preps sequence obtained		64
Sequence rejected (cross contamination of adjacent wells)	1	
Total insertions rejected	1	
Independent insertions for which location was obtained		63
Number of insertion locations		62
Number inserted clockwise	33	
Number inserted counterclockwise	30	
Aberrant insertions		
Number of insertion plasmids with structural aberrations	1	
This was a deletion far from the insertion		
Number of structural aberrations associated with insertion site	0	
Number of insertions with disagreement in 5 bp duplication	2	

These were:

G-->A transition mutation in one copy with respect to target plasmid sequence

G-->T transversion mutation in one copy with respect to target plasmid sequence.

b. Analysis of the distribution of insertions among sequences and intervals.

For the purpose of obtaining maximum sequence from an unknown target, it is desirable that the insertions be distributed as randomly as possible with respect to regions of sequence and with respect to specific sequences. The summary of Table 4

already suggests a very random process, since 63 independent insertions hit 62 different locations, i.e. no hotspots for insertion were identified. For comparison, relaxed-specificity derivative of Tn10 (ATS2, examined with *in vivo* insertions into the *lac* operon) hit 23 sites with 50 insertions.

Primary data for further analysis below is found in Table 5, which gives the location of all the insertions, their orientation with respect to the target plasmid, and the sequence immediately adjacent to the insertion (the five bp sequence duplicated by the insertion mechanism) in a uniform frame of reference.

Table 5: Insertion locations and associated 5 bp duplication

	Directions			Sequence at position relative to Tn7R					orient-ation
Isolate	Sequence obtained	Location	Insert name	1	2	3	4	5	(Tn7 R clock-wise= +)
1	1	6464	A5	A	G	C	T	C	-
2	2	8428	A6	C	T	G	G	T	-
3	1	8349	A7	C	C	T	G	A	+
4	2	5515	A8	T	A	A	C	T	+
5	2	7822	A9	C	C	C	G	C	+
6	2	365	A10	T	C	A	A	C	+
7	2	5695	A11	T	C	A	C	G	-
9	2	2500	B1	G	G	A	T	G	+
10	1	8286	B2	C	T	T	C	C	+
11	1	2764	B3	C	T	T	T	A	+
12	2	6953	B4	C	G	A	G	G	+
13	2	3414	B5	C	T	T	T	G	+
14	1	3139	B6	T	C	G	T	T	-
15	1	3208	B7	G	C	A	C	T	-
16	2	4208	B8	A	G	A	G	C	-
17	2	3671	B9	G	T	T	T	A	+
18	2	5563	B10	C	C	A	A	C	-
19	1	3539	B11	G	C	T	T	C	+
20	2	3803	B12	A	T	T	C	C	-

21	2	8474	C1	C	C	G	C	C	+
22	2	5661	C2	A	T	G	A	T	+
23	2	7693	C3	C	G	C	G	T	-
24	2	3205	C4	T	C	T	T	C	-
25	2	1650	C5	C	C	T	A	T	-
26	2	8020	C6	G	C	C	G	G	-
27	2	2566	C7	A	T	T	T	T	+
29	2	2275	C9	G	C	C	C	A	+
30	1	6368	C10	G	C	T	A	T	-
32	2	2629	C12	T	A	T	A	C	+
33	2	5988	D1	G	G	C	G	A	+
34	2	3499	D2	A	T	G	T	A	-
35	2	3933	D3	T	T	G	A	T	-
36	2	6077	D4	G	T	T	G	T	+
37	2	6756	D5	T	T	G	A	G	-
38	2	5563	D6	G	T	T	G	G	+
38	2	8224	D7	G	G	A	G	G	-
40	2	3123	D8	C	A	A	A	T	-
41	1	2746	D9	A	A	A	A	C	-
42	1	1646	D10	C	G	A	G	A	+
43	1	5678	D11	A	T	G	T	G	+
44	2	7406	D12	T	G	C	A	T	+
45	2	1744	E1	G	C	C	A	T	-
46	2	3584	E2	T	A	G	G	T	+
47	2	2112	E3	C	C	T	A	C	+
48	2	4205	E4	G	C	A	G	C	-
49	1	2708	E5	G	C	G	G	T	+
50	2	7828	E6	A	C	A	G	A	+
52	2	3873	E8	A	G	T	C	T	-
53	2	3591	E9	C	A	T	G	C	-
56	2	5550	E12	A	T	C	G	C	-
57	2	2702	F1	T	T	C	A	C	+
61	2	4490	F5	G	T	T	A	A	-
62	2	5811	F6	A	C	G	C	G	+
63	2	2024	F7	A	C	T	G	T	-
64	2	1479	F8	A	T	C	G	T	-
66	2	5675	F10	T	T	T	A	T	+
67	2	5208	F11	A	T	A	A	A	+
68	2	6020	F12	G	G	T	A	A	+
69	2	6264	G1	G	A	G	T	A	+
70	2	3881	G2	A	T	T	T	G	-
72	2	2891	G4	A	T	T	C	G	-
74	2	1681	G6	A	C	T	C	T	-
76	2	5315	G7	A	A	T	A	C	+

Table 5 legend:

Isolate: Number of the colony

Directions sequenced: 1= only one direction from the insertion; 2= both directions

Position: coordinate on pER183 (SEQ ID NO:5) top strand of the first base of the 5 bp duplication

Insert name: accession number in notebook

Sequence at position #: position 1 is the base immediately adjacent to Tn7R top strand (i.e. it can be either the top or the bottom strand of pER183 (SEQ ID NO:5)); position #2 is the next but one to Tn7R; and so forth.

Orientation: of the insertion relative to the top strand of pER183 (SEQ ID NO:5). +, Tn7R is to the right of Tn7L when displayed on the top strand of pER183 (SEQ ID NO:5). -, Tn7R is to the left of Tn7L.

1. Distribution of insertions fits the Poisson distribution

1. These insertions are randomly distributed as judged by the fit of the interval distribution to the distribution predicted by a Poisson process.

The Poisson distribution gives the probability of observing exactly X_i events (insertions) in a unit (interval) when the average number of events per unit is μ (from Zar, J.H. *Biostatistical Analysis* Prentice-Hall, Englewood Cliffs, NJ 1974 p.301).

$$P(X_1) = \frac{\mu^{Xi} e^{-\mu}}{X_i!} \qquad \text{eq 1}$$

Where

X_i= exactly X_i insertions per interval

μ = average number of insertions per interval

Let

X_i= number of insertions in a 100 bp interval

f_i= Observed number of 100 bp intervals with X_i insertions/interval

n = number of 100 bp intervals in the set (=73)

$\mu = \Sigma f_i X_i / \Sigma f_i = 63/73$

$P_{(xi)}$ = probability of finding X_i insertions in a 100 bp interval (from the Poisson distribution,. eq 1)

$F_i = P_{(xi)}$ n = expected number of intervals with i insertions.

From the data in Table 5 and eq 1 we can construct the following comparison of expected and observed data:

Table 6. Observed and expected distribution of insertions in 100 bp intervals

Insertions per interval	Observed intervals with X_I insertions	Probability of X_I insertions per interval	Expected number of intervals with X_I insertions
X_i	f_i	$P(X_i)$	F_i
0	34	0.42189	30.80
1	24	0.35410	26.58
2	9	0.15711	11.47
3	3	0.04520	3.299
4	3	0.00975	0.712

These distributions are illustrated in Figure 12, where fi= observed distribution, Fi = expected distribution. The fit looks good to the eye.

b. Statistical test of fit between observed and expected distributions

To test whether the observed and expected distribution are statistically indistinguishable, we used a Chi-square test for goodness of fit (from Zar, J.H. *Biostatistical Analysis* Prentice-Hall, Englewood Cliffs, NJ 1974 p. 303). For this purpose we pool the tail of the distribution so that no expected number is less than 4. Rewriting Table 6, we obtain

Table 7. Chi-square test of goodness of fit to a random distribution

Insertions per Interval	Observed intervals with X_I insertions	Expected number of intervals with X_i insertions	$(f_iF_i)^2$ ------- F_i
X_i	f_i	F_i	Chi-square
0	34	30.80	0.3329
1	24	26.58	0.2504
2	9	11.47	0.5315
≥3	6	4.1541	0.8209
		Sum	1.944

The null hypothesis is that the observed distribution was drawn from a Poisson distributed population. For two degrees of freedom this sum of chi-square values gives a probability that this is the case of $0.25<p<0.5$. The null hypothesis is not rejected.

In sum, the eye (part a, Fig. 12) and a statistical test (Table 7 and following) agree that the distribution of insertions in intervals along the DNA is random.

ii. Analysis of the base composition of insertion sites.

Site preference of $TnsABC_{127}$ for insertion of miniTn7 into pER183

Certain bases are preferred at some positions in the five-base insertion site duplication, as shown in a histogram of base incidence versus position in the site (Figure 13), taken from the data in Table 5. In collating the data for this histogram, the five duplicated bases were assigned position numbers relative to Tn7R; position one is the base immediately adjacent to Tn7R when the sequence is displayed with Tn7R on the right and Tn7L on the left. The orientation of the transposon relative to the target sequence during target choice is thus

controlled for: the target is displayed in the same way relative to the transposon for all insertion sites.

A model for a preferred site was formulated: NYTRN. The elements of this site were tested for statistical significance individually and collectively by chi-square analysis (Table 8). The null hypothesis was that sites were drawn randomly from the universe of sequence defined by the sequence of pER183 (SEQ ID NO.5) after deleting sequence subject to selection (bp 1-250 and 2481-2509, CamR; and 581-1400, replication origin). Expected frequencies of the four bases, of purines and pyrimidines, and of trinucleotides were derived from frequencies obtained for pER183 (SEQ ID NO:5)-condensed by the GCG program COMPOSITION.

Table 8. Chi-square tests (tests that differ from random expectation (p<0.05) in bold)

Four bases individually, all sites collectively (315 bp experimental, 7410 bp control)

Base	Expected	Observed	Chisquare	probability
A	78.4	73	.372	
C	74.3	77	.981	
G	76.2	72	.232	
T	85.7	93	.622	
			2.21	0.5<p<0.75

Four bases individually, each position individually (63 bp experimental, 7810 bp control)

Position 1

A	15.7	19	0.694	
C	14.9	15	0.00066	
G	15.2	17	0.213	
T	17.1	12	1.52	
			2.43	0.25<p<0.5

	Position 2				
	A	15.7	8	3.8	
	C	14.9	22	3.38	
	G	15.2	11	1.16	
	T	17.1	22	1.4	
				9.74	**0.01<p<0.025**
	Position 3				
	A	15.7	15	0.031	
	C	14.9	11	1.02	
	G	15.2	12	0.674	
	T	17.1	25	3.65	
				5.37	0.1<p<0.25
	Position 4				
	A	15.7	19	0.693	
	C	14.9	11	1.02	
	G	15.2	20	1.52	
	T	17.1	13	0.983	
				4.21	0.1<p<0.25
	Position 5				
	A	15.7	12	0.872	
	C	14.9	18	0.645	
	G	15.2	12	0.674	
	T	17.1	21	0.889	
				3.08	0.25<p<0.5

Purines and Pyrimidines, each position individually (63 bp experimental, 7410 bp control)

Base	Expected	Observed	Chisquare	probability
Position 1				
R	30.9	36	0.842	
Y	32.1	27	0.810	
			1.65	0.1<p<0.25
Position 2				
R	30.9	19	4.58	
Y	32.1	44	4.41	
			8.99	**0.001<p<0.005**
Position 3				
R	30.9	27	0.49	
Y	32.1	36	0.422	
			0.914	0.25<p<0.5
Position 4				
R	30.9	39	2.12	
Y	32.1	24	2.04	
			4.17	**0.025<p<0.05**
Position 5				
R	30.9	24	1.54	
Y	32.1	39	1.48	
			3.10	0.05<p<0.1
T or not-T, position 3				
T	17.16	25	3.58	
not-T	45.84	38	1.34	
			4.92	**0.025<p<0.05**

Triplets, positions 234 (63 experimental triplets, 7408 control triplets to determine expectation)

All triplets

Triplet	Expected	Observed	Chisquare	probability
YNR	15.96	25	7.54	
RNY	15.97	5	5.12	
RNR	14.98	14	0.064	
YNY	16.07	19	0.534	
			13.25	**0.001<p<0.005**

Specific triplets, positions 234

YNR	16	25	5.06	
Not YNR	47	38	1.7	
			6.78	**0.005<p<0.01**
RNY	16	5	7.56	
Not RNY	47	59	3.06	
			10.62	**0.001<p<0.005**
YTR	3.93	10	9.38	
not YTR	59.07	53	0.623	
			10.0	**0.001<p<0.005**

Pairing between position 2 and 4 (GNC, CNG, ANT, TNA)

Paired	16.95	16	0.053	
Not paired	46.05	47	0.0196	
			0.073	0.75<p<0.9

Preference for this site was statistically significant ($p<0.005$), and preference for each of its parts was also significant ($p<0.05$). However, the preference is not particularly strong, in that representation of the site was only 2.5-fold more frequent in insertion sites than expected from the composition of the plasmid; and 53 out of 63 sites do not fit the consensus. Each preferred position contributes independently to the overall preference, since multiplying together the overrepresentation of each position yields the overrrepresentation of the site as a whole (Table 9).

Table 9: overrepresentation of preferred bases in Tn7 insertion sites

Position	preference	expected	observed	Fold overrepresentation (Obs/Exp)
2	Y	32.1	44	1.37
3	T	17.6	25	1.42
4	R	30.9	39	1.26
product ((O/E)2 x (O/E)3 x (O/E)4)				2.46
triplet	YTR	3.93	10	2.54

We conclude that insertion mediated by $TnsABC_{127}$ is extremely random, with only a slight preference for sites of the form NYTRN (SEQ ID NO:15).

I. Recipes.

1. 100 X DNA buffer per liter

Tris Base 121.1 g

Dissolve in 700 ml

4 M HCl ~90ml

Bring pH to 7.4

Na_2EDTA	37.2
NaCl	29.22 g

Make up to ~950 ml
adjust pH
Make up to 1 L
Aliquot, autoclave

2. 1 X DNA buffer

100 x DNA buffer	1	ml
dH_2O, sterile	100	ml

3. 10 X Fi/L (Fill-in, ligation) buffer

10 X ligase buffer	1500	µl

New England BioLabs, 32 Tozer Road, Beverly, Massachusetts, 01915

100 mM dATP	3.75 µl

New England BioLabs, 32 Tozer Road, Beverly, Massachusetts, 01915

100 mM dCTP	3.75 µl

New England BioLabs, 32 Tozer Road, Beverly, Massachusetts, 01915

100 mM dGTP	3.75 µl

New England BioLabs, 32 Tozer Road, Beverly, Massachusetts, 01915

100 mM dTTP	3.75 µl

New England BioLabs, 32 Tozer Road, Beverly, Massachusetts, 01915

4. TE

1 M Tris pH 8.0	1 ml
0.5 M EDTA pH 8.0	0.2 ml
dH_2O	to 100 ml

Filter sterilise

Example 5

A convenient method for stopping a transposon insertion reaction

In order to use DNA molecules with transposon insertions, they must be recovered *in vivo*. It is most convenient to be able to do this without the labor and losses associated with extraction with organic solvents and alcohol precipitation. Prior art has suggested, however, that transposition reaction products formed during *in vitro* insertion experiments are DNA: protein complexes that are extremely stable; evidence suggests that a chaperone-like activity is required for disruption of these products. Accordingly, organic extraction was deemed required for satisfactory disruption of the complexes.

This example demonstrates that heat inactivation at 75°C is adequate for disrupting these complexes or at least for putting them into a form that can be introduced into the cell by chemical transformation.

Section A. MATERIALS

A1) PROTEINS

TnsA 30 µg/ml in 10% glycerol

TnsB 20 µg/ml in 25% glycerol

$TnsC_{127}$ 100 µg/ml in 10% glycerol

Keep stored at -70°C. Sufficient protein for 10 reactions is provided. At the time of use, keep frozen on dry ice until ready to add to the reaction, and keep on dry ice until returned to the freezer.

A2) BUFFER CONSTITUENTS

HEPES	0.25 M pH 8.1
Tris[Cl]	0.25 M pH 7.6

BSA	10 mg/ml
tRNA	50 µg/ml
DTT	1 M
ATP	100 mM
MgAcetate	375 mM

A3) TRANSPOSON DONOR PLASMID

100 µg/ml

This is as described for Example 3.

A4) TARGET PLASMID

pLITMUS28 400 µg/ml

A5) OTHER

Millicue water

Heat block, 30^0C

1.5 ml microtubes.

A6) FOR STOPPING THE REACTION

when using chemically competent cells

Water bath or heat block, 75°C.

Water bath or heat block, 65°C

Distilled phenol equilibrated with TE

Chloroform equilibrated with TE

EtOH for precipitation

NaCl 3 M

Water or 1 mM Tris pH 8

CHEMICALLY COMPETENT TRANSFORMABLE CELLS:

In this example, we show the use of

Chemically competent ER 1821 (2 x 10^7 transformants/μg of LITMUS

Chemically competent ER2502 (6 x 10^6 transformants/μg of LITMUS

prepared as in Example 3

A8) MEDIA

Rich Broth and Rich Agar (Kan, Amp) prepared as in Example 3.

Section B. Tn7 *in vitro* transposition reaction protocol

1. Experiment 1. Four stop treatments

Reactions were carried out as in Example 3, using quadruplicate samples for each of four treatments. At step 12, one of these treatments was substituted. For transformation, ER2502 was used.

Treatment 1: No treatment.

Treatment 2: Heat treatment at 65°C for 20 min

Treatment 3: Heat treatment at 75°C for 10 min.

Treatment 4: Phenol extraction once, chloroform extraction once, ethanol precipitation once, resuspension in original volume of TE.

The results of this experiment are given in Table 9 below and illustrated in Figure 14

Table 10: Transformants obtained per 1/50th volume of transposition reaction

Replicate

	#1	#2	#3	#4
nothing	1	0	0	0
65C 20'	1	0	0	0
75C 10'	32	24	22	10
phenol+pptn	30	23	20	17

2. Experiment 2. Three stop treatments

Reactions were carried out as in Example 3, using duplicate samples for each of three stop treatments, for two aliquots of TnsB, and for three volumes of TnsB. At step 12, one of the stop treatments was substituted. For transformation, ER1821 was used.

Treatment 1: Heat treatment at 75°C for 10 min.

Treatment 2: Ethanol precipitation only, resuspension in original volume of TE

Treatment 3: Heat treatment at 65°C for 20 min

The results of this experiment are given in Table 9 below and illustrated in Fig. 15

Table 9. Transformants obtained per 1/50th volume of transposition reaction, three stop treatments.

	TnsB		75C 10 min		EtOH pptd		65C 20 min.	
	Aliquot	Volume (µl)	#1	#2	#1	#2	#1	#2
	1	1	158	8	13	3	15	10
	1	1.5	186	0	16	0	3	0
	1	2	178	170	13	13	30	16
	1	3	454	366	47	21	11	8
	2	1	324	140	21	3	9	2
	2	1.5	506	462	58	44	25	25
	2	2	1220	1102	88	37	54	18
	2	3	1802	1690	129	126	37	14

These two experiments demonstrate that heat treatment at 75°C for 10 min is an adequate method of stopping the transposition reaction and gives as many transformants as treatment with phenol, chloroform and ethanol precipitation; whereas no treatment, ethanol precipitation alone, and heat treatment at 65°C for 20 min is inadequate, giving no transformants or a greatly reduced number of transformants.

Example 6

Storing three components of Tn7 transposase together

Convenient routine use of *in vitro* transposition as a method in molecular biology would be facilitated if the protein components of the reaction could be stored in a single tube. In this way, variability in volume measurement from one experiment to another would be minimized, time and labor would be saved, and reproducibility enhanced. The $TnsABC_{127}$ transposition reaction described in the foregoing examples involves the addition of three different protein components.

This example demonstrates that these three protein components of the reaction can be mixed and stored together without interfering with the efficiency of the transposition reaction.

Section A. MATERIALS

A1) INDIVIDUAL PROTEINS

TnsA 30 µg/ml in 10% glycerol

TnsB 20 µg/ml in 50% glycerol

$TnsC_{127}$ 100 µg/ml in 50% glycerol

Keep stored at -70^{0}C.

A2) MIXED PROTEINS, COMPRISING

TnsA 7.36 µg/ml

TnsB 11.3 µg/ml

$TnsC_{127}$ 18.9 µg/ml

in 40% glycerol

A2a) Keep stored at -70□C, or

A2b) Keep stored at -20□C

A3) OTHER COMPONENTS

These are as in example 1, parts A and B; including chemically competent ER2502 (6 x 10^6 transformants/μg of LITMUS) prepared as in example 1.

Section B. Tn*7* *IN VITRO* TRANSPOSITION REACTION PROTOCOL

B1. Reaction volume = 100 μl

B2. Experimental variations (2 experiments are shown, reactions were carried out in quadruplicate).

Tube 1 Proteins of A1 added individually at step 6 below in a total volume of 5.3 μl

Tube 2 Mixture of A2a added together at step 6 below in a total volume of 5.3 μl

Tube 3 Mixture of A2b added together at step 6 below in a total volume of 5.3 μl (Experiment 2 only)

B3. Make up a mix as in Example 1, section C

B4. Dispense mix of step 3 as in Example 1, section C

B5. Add target DNA as in Example 1, section C. In this example, this is pLITMUS28, 1 μl

B6. Add to each tube

		Tube 1	Tube 2	Tube 3
TnsA	1.3	μl (40 ng)		
TnsB	3	μl (20 ng)		
$TnsC_{127}$	1	μl (100 ng)		
$TnsABC_{127}$	0		5.3 μl (39ng A, 59.9 ng B 100.2 ng C_{127})	5.3 μl (39ng A, 59.9 ng B 100.2 ng C_{127})

B7. Add 1µl donor DNA (0.1 µg pMCB40) as in example 1C

B8. Incubate 10 minutes at 30°C (assembly reaction) as in example 1C

B9. Add 4 µl MgAc (375mM) to each tube as in example 1C

B10. Incubate 1 hour 30°C (transposition reaction) as in example 1C

B11. Heat Inactivate 75°C 10 minutes

B12. Transform using chemically competent cells, as in example 1.

In this example, selective medium was RB Kan (20 µg/ml) Amp (100 µg/ml). Competent cells were ER2502, chemically competent (Example 1, Section D1).

C. Transformation result:

Experiment 1: Proteins were stored individually at -70°C or as a mixture at -70°C (A2a). In this experiment, the proteins in both treatments had suffered the same number of freeze-thaw cycles. 10 µl of each 100 µl reaction was transformed, and 100 µl of the 500 µl outgrowth culture was plated.

Table 10: Transformants obtained per 1/50th volume of transposition reaction, transposition proteins added as a mixture or individually. Result is displayed in Figure 16.

	Replicate					
Storage	#1	#2	#3	#4	Average	avg per reaction
Individually	27	62	59	41	47	2350
As a mixture	47	68	60	23	49	2450

Experiment 2: Proteins were stored individually at -70°C, as a mixture at -70°C (A2a material) or as a mixture at -20°C (A2b material). In this experiment, the proteins stored

individually had suffered more freeze-thaw cycles than those stored together. 10 µl of each 100 µl reaction was transformed, and 100 µl of the 500 µl outgrowth culture was plated.

Table 11: Transformants obtained per 1/50th volume of transposition reaction, transposition proteins added as a mixture or individually following storage at -20°C or -70°C. Result is displayed in Figure 17.

	Replicate					
Storage	#1	#2	#3	#4	average	avg per reaction
Individually	13	38	17	22	22	1100
As a mixture, -70°C, (A2a)	167	173	117	218	168	8400
As a mixture, -20°C, (A2b)	179	125	219	199	180	9000

These two experiments demonstrate that the three Tns proteins can be stored together. The difference in experiment 2 between individual storage and storage together may be attributed to the number of freeze-thaw cycles.

Bibliography

1. Craig NL. 1997. In Ann Rev Biochem. ed. pp. 437-74. Palo Alto: *Annual Reviews Inc.*
2. Kleckner N. 1989. In Mobile DNA. ed. Berg D and Howe M, pp. 227-68. Washington, D.C.: *American Society for Microbiology.*
3. van Luenen HGAM, *et al.* 1994. *Nucleic Acids Res. 22*:262-94. Rosenzweig B, *et al.* 1983. *Nucleic Acids Res. 11*:4201-10
5. Mori I, *et al.* 1988. *Proc Natl Acad Sci USA. 85*:861-4
6. Eide D, *et al.* 1988. *Mol Cell Biol. 8*:737-46
7. Mizuuchi M, *et al.* 1993. *Cold Spring Harbor Symp. Quant. Biol. 58*:515-23

8. Berg DE. 1989. In Mobile DNA. ed. Berg D and Howe M, pp. 185-210. Washington, D.C.: *American Society for Microbiology.*
9. Kirchner J, *et al.* 1995. *Science. 267*:1443-4
10. Rogers M, *et al.* 1986. *Mol. Gen. Genet. 205*:550-6
11. Waddell CS, *et al.* 1988. *Genes Dev. 2*:137-49
12. Sarnovsky R, *et al.* 1996. *EMBO J. 15*:6348-61
13. Kubo KM, *et al.* 1990. *J. Bacteriol. 172*:2774-8
14. Wolkow CA, *et al.* 1996. *Genes Dev. 10*:2145-57
15. Hauer B, *et al.* 1984. *Mol Gen Genet. 194*:149-58
16. Arciszewska LK, *et al.* 1989. *J. Mol. Biol. 207*:35-52
17. Lee C-H, *et al.* 1983. *Proc Natl Acad Sci USA. 80*:6765-9
18. Reyes I, *et al.* 1987. *Plasmid. 18*:183-92
19. Adzuma K, *et al.* 1988. *Cell. 53*:257-66
20. Sakai J, *et al.* 1995. *EMBO J. 14*:4374-83
21. Kleckner N, *et al.* 1996. *Curr Top Microbiol Immunol. 204*:49-82
22. Bainton R, *et al.* 1991. *Cell. 65*:805-16
23. Bainton RJ, *et al.* 1993. *Cell. 72*:931-43
24. Gamas P, *et al.* 1992. *Nucl. Acids Res. 20*:2525-32
25. Stellwagen A, *et al.* 1997. *Genetics. 145*:573-85
26. Stellwagen A, *et al.* 1997. *EMBO J.* (in press):
27. Sankar P, *et al.* 1993. *J.Bacteriol. 175*:5145-52
28. Devine SE, *et al.* 1994. *Nucleic Acids Res. 22*:3765-72
29. Pryciak PM, *et al.* 1992. *Proc Natl Acad Sci USA. 89*:9237-41
30. Pryciak PM, *et al.* 1992. *Cell. 69*:769-80
31. Pryciak PM, *et al.* 1992. *Embo J. 11*:291-303
32. Singh IR, *et al.* 1997. *Proc Natl Acad Sci U S A. 94*:1304-09
33. Kholodii G, *et al.* 1995. *Mol. Microbiol. 17*:1189-200
34. Radstrom P, *et al.* 1994. *J. Bacteriol. 176*:3257-68
35. Reimmann C, *et al.* 1989. *Mol Gen Genet. 215*:416-24

36. Rowland S-J, *et al.* 1990. *Mol. Microbiol. 4*:961-75

37. Walker, *et al.*, eds. 1983. *Techniques in Molecular Biology*. New York: MacMillan Publishing Company

38. Kunkel. 1985. *Proc Natl Acad Sci USA. 82*:488-92

39. Kunkel, *et al.* 1987. *Methods Enzymol. 154*:367-82

40. Sambrook J, *et al.*, eds. 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Press

41. Dayhoff, *et al.*, eds. 1978. Washington, D.C.: *Natl. Biomed. Res. Found.*

42. Miller JH. 1972. *In Experiments in Molecular Genetics*. ed. pp. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

43. Elespuru RK, *et al.* 1979. *Environ Mutagen. 1*:65-78

44. Yarmolinsky MB, *et al.* 1983. *Mol Gen Genet. 192*:140-8

45. Haniford DB, *et al.* 1989. *Cell. 59*:385-94

46. McKown RL, *et al.* 1987. *Proc.Natl.Acad.Sci.USA. 84*:7807-11

47. McKown RL, *et al.* 1988. *J. Bacteriol. 170*:352-8

48. Johnson RC, *et al.* 1984. *Genetics. 9-18*

49. Rose MD, *et al.*, eds. 1990. *Methods in Yeast Genetics: A Laboratory Course Manual.* Cold Spring Harbor: Cold Spring Harbor Laboratory

50. Hughes O. 1993. *Host Components of Tn7 Transposition.*

51. Huisman O, *et al.* 1987. *Genetics. 116*:185-9

52. DeBoy R, *et al.* 1996. *J. Bacteriol. 178*:6184-91

53. Flores C, *et al.* 1990. *Nucl. Acids Res. 18*:901-11

54. Walker JE, *et al.* 1984. *Biochem J. 224*:799-815

55. Saraste M, *et al.* 1990. *Trends Biochem Sci. 15*:430-4

56. Sancar A, *et al.* 1993. *Science. 259*:1415-20

57. Chaconas G, *et al.* 1985. *J Biol Chem. 260*:2662-9

58. Faelen M, *et al.* 1978. *Nature. 271*:580-2

59. O'Day KJ, *et al.* 1978. *In Microbiology*. ed. Schlessinger D, pp. 48-51. Washington, D.C.: American Society of Microbiology.

60. Surette MG, *et al.* 1987. *Cell. 49*:254-62

61. Craigie R, *et al.* 1987. *Cell. 51*:493-501

62. Koonin EV. 1992. *Nucleic Acids Res. 20*:1997

63. Gary PA, *et al.* 1996. *J. Mol. Biol. 257*:301-16

64. Gwinn ML, *et al.* 1997. *J. Bacteriology. 179*:7315-20

65. Bender J, *et al.* 1992. *EMBO J. 11*:741-50

66. Lichtenstein C, *et al.* 1982. Unique insertion site of Tn7 in E coli chromosome. *297*:601-3#001#